THE UNIVERSITY
of EDINBURGH

Investigating the role of chromatin modifications in CRISPR/Cas9 gene editing.

Eirini – Margarita Kallimasioti – Pazi

Supervisor: Dr Andrew J Wood

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

MRC Human Genetics Unit

2020

Abstract

Precisely positioned nucleosomes and heterochromatin have been shown to impede CRISPR/Cas9 editing efficiency. Conversely, Cas9 can open previously inaccessible regions of DNA, and transcriptionally silent targets can usually be edited without difficulty. In order to address this paradox, we have developed a method that exploits the allele-specific chromatin status of imprinted genes to characterise the impact of chromatin modifications on targeted mutagenesis. Cas9 was targeted to imprinted CpG islands in F1 hybrid mouse embryonic stem cells, and then allele-specific mutation patterns were characterised following high throughput amplicon sequencing. Using this novel system, we discovered that heterochromatin can impede mutagenesis with CRISPR/Cas9, but to a degree that depends on other key experimental parameters. Mutagenesis was impeded when Cas9 exposure was brief and when intracellular expression of Cas9 was low, but the consequences of chromatin modifications were minimal following prolonged exposure. The presence of mismatches between single guide RNA and genomic target sequence disproportionately reduced mutagenesis within heterochromatin for some specific combinations of mismatches. This suggests that Cas9 proofreading activity, and hence off-target mutagenesis, may be impacted by chromatin state and merits further investigation. Lastly, no effects of chromatin modifications on the outcome of DNA repair were detected, with similar efficiencies of homology-directed repair (HDR) and non-homologous end joining (NHEJ) on maternal and paternal alleles. Upon further characterisation we discovered that the majority of types of insertions and deletions generated by NHEJ shared similar frequencies between the two alleles. Combined, my data show that heterochromatin imposes a permeable barrier that influences the kinetics, but not the endpoint, of CRISPR/Cas9 genome editing and suggest that therapeutic applications involving low-level Cas9 exposure will be particularly affected by chromatin status.

Lay Summary

One of the biggest advances in biology during the early part of the 21st century has been the development of methods to precisely modify DNA sequence in living cells. This process, termed ‘genome editing’, is achieved by using special tools such as CRISPR/Cas9. CRISPR/Cas9 is very efficient, cost-effective and can be easily adjusted to modify any desired DNA sequence. Editing is achieved by creating a break in the DNA through positioning the Cas9 protein on DNA sequence of interest. An externally provided DNA template can be used to incorporate any desired sequence change while the break is undergoing repair.

The complete DNA sequence in every living cell is packaged with the aid of proteins. Consequently, DNA packaging determines which DNA sequences can be hidden or exposed at any time. Emerging evidence suggests that DNA packaging may pose a natural barrier to the action of Cas9. However, it is unknown whether additional factors play a role during this inhibition. In addition, it is currently unclear if DNA packaging affects the incorporation of desired modifications during genome editing with CRISPR/Cas9.

During my PhD, I established a new experimental system to study how different modes of DNA packaging affect the ability of CRISPR/Cas9 to edit mammalian DNA sequences. This led to the discovery of other factors that influence how DNA packaging affects CRISPR/Cas9 activity. It also became evident that how the target DNA was packaged before it is broken by CRISPR/Cas9 does not affect the sequence changes that arise following repair. These findings help us to understand how CRISPR/Cas9 works within mammalian cells and could promote its safe use in the clinic for novel therapeutic purposes.

Acknowledgements

First of all, I would like to thank my primary supervisor, Andrew Wood, for providing me with inspiration and support throughout the years of my doctorate study. If I had to distinguish the time he helped me most, it would be that fateful day in Keystone, Colorado. For his supervision, trust and open-mindedness I will remain grateful. I would also like to thank the members of my PhD committee, Wendy Bickmore, Abdenour Soufi and Martin Taylor for their support and encouragement. Especially my assistant supervisor, Wendy Bickmore, for providing some of the most constructive advice on how to move forward in my PhD.

Secondly, I would like to acknowledge all the present and past members of the Wood lab for day-to-day supervision and the amazing working atmosphere. Keerthi Chathoth and Gillian Taylor and Martijn Kelder deserve a special place in my acknowledgements for helping me optimise the main system I used throughout my project. Jess, Lewis and Jenny, thank you for all the lovely memories in the lab. I would also like to express my gratitude to Alison Meynert and Tracy Ballinger for helping me with Next Generation Sequencing data analys. In addition, I would like to thank Heidi Mjøseng for her help, especially during the busiest and most tiresome period of my PhD. I will never forget Ryu-suke Nozawa, a true friend who also taught me the true meaning of efficiency in the lab.

Finally, I would like to thank my two best friends Kristina Kovacikova and Emma Katarina Jönsson (or simply Kris and Kat). Thank you for all the laughs, trips, advice and motivation to keep fighting whatever the circumstances. Special thanks to my parents, Dimitrios Kallimasiotis-Pazis and Anastasia Barmpagianni, for being there for me at any time of need and encouraging me to keep my head up and go on. Last but not least, I would like to acknowledge my partner, Emmanouil Athanasopoulos. Thank you for all your love and support, despite the long distance between us.

Declaration of originality

I declare that this thesis was written by myself. Unless otherwise stated and acknowledged in the text, the work presented here is my own. This work has only been submitted for the degree of Doctor of Philosophy at the University of Edinburgh.

The data from Chapters 3, 4 and 5 (excluding Chapter 4.5, Chapter 5.3 and Chapter 5.5) have also recently been published in the following peer-reviewed research article:

Kallimasioti-Pazi, E. M. *et al.* (2018) 'Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair', *PLoS Biology*. doi: 10.1101/267690

Eirini – Margarita Kallimasioti – Pazi

Edinburgh, 2020

List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Actb	Beta actin
alt-EJ	Alternative end joining
ANOVA	Analysis of variance
ATAC-Seq	Assay for Transposase-Accessible Chromatin using sequencing
BLM	Bloom helicase
Cas9	CRISPR-associated protein 9
Cas9n	Cas9 nickase
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation followed by sequencing
c-NHEJ	Cassical non-homologous end joining
CRISPR	Clustered regularly interspaced short palindromic repeats
dCas9	Catalytically dead CRISPR-associated protein 9
DHS	DNase I hypersensitive site
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	Double-strand break
FA	Fanconi Anemia pathway

FACS	Fluorescence-activated cell sorting
FRET	Förster resonance energy transfer
HDR	Homology-directed repair
hESCs	Human embryonic stem cells
hHSPCs	Human hematopoietic stem/progenitor cells
HP1	Heterochromatin protein 1
HR	Homologous recombination
IAP	Intracisternal A particle
ICRs	Imprinting control regions
IF	Immunofluorescence
Igf2	Insulin-like growth factor 2
Igf2r	Insulin-like growth factor 2 receptor
InDels	Insertions and deletions
LEDGF	Lens epithelium–derived growth factor
lncRNAs	Long non-coding RNAs
MBDs	Methyl-CpG-binding domains
mESCs	Mouse embryonic stem cells
MMEJ	Microhomology-mediated end joining
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PRC2	Polycomb repressive complex 2
qPCR	Quantitative polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNPs	Ribonucleoprotein particles
RPA	Replication protein A
SDSA	Synthesis-dependent strand annealing
sgRNA	single guide RNA
SIM	structured illumination microscopy
SNP	single nucleotide polymorphism
SSA	Single-strand annealing
ssODN	single-stranded oligodeoxynucleotide
TALEs	Transcription activator-like effectors
tracrRNA	trans-activating crRNA
TSA	Trichostatin A
ZFN	Zinc finger nuclease

List of figures

Table of Contents

Chapter 1

Introduction

1.1 Genome editing technology in mammalian cells

1.1.1 Early approaches and gene targeting

Genome editing is a valuable method for introducing desirable DNA sequence modifications, such as specific insertions, deletions and substitutions, to genomic locations of interest. By modifying the DNA sequence scientists can study the function of protein coding genes and non-coding regulatory elements, as well as develop model systems to systematically characterise disease phenotypes (Tan *et al.*, 2012).

Before the development of the currently used approaches for generating targeted modifications, mutagenesis was performed in a random manner using chemical substances, or other mutagenic sources such as irradiation, and screening the resulting individual phenotypes (Stanford, Cohn and Cordes, 2001). In particular, such random mutagenesis strategies were pivotal to dissect biological functions encoded within the genomes of classical model organisms, such as bacteria, yeast, nematode worms, the fruit fly and later also in mice (Muller, 1928; Brenner, 1974; Capecchi, 2001).

This approach was succeeded by gene targeting, a method for generating specific modifications with the aid of homologous recombination (HR), which is an endogenous DNA repair mechanism (Smithies *et al.*, 1985; Doetschman *et al.*, 1987; Thomas and Capecchi, 1987). Basic gene targeting is achieved through the integration of an externally provided DNA template that contains a selectable marker gene and the desirable edit, flanked by regions of homology to the genomic target site (Tokunaga, Anai and Hanada, 2016). The advantage of gene targeting over preceding approaches is its ability to precisely target genomic sequences of interest, instead of a large number of random undesirable sites. In addition, the requirement of HR allows specific sequences of interest to be inserted at endogenous loci while minimising the chances of random integration.

However, gene targeting also possesses several caveats. The size of mammalian genomes, being larger than the genomes of other unicellular eukaryotic species like yeast, poses certain difficulties with obtaining desirable changes. Despite the fact that gene targeting was widely used to generate numerous mouse models to study the function of mammalian genes (Capecchi, 2005) the efficiency of the technique remained low (Lin, Sperle and Sternberg, 1985; Thomas, Folger and Capecchi, 1986). In order to obtain precisely engineered cells researchers would screen many clones, resulting in a process that requires time and creates significant technical challenges. Furthermore, gene targeting was only efficient within a small selection of cell lines, including mouse embryonic stem cells (mESCs) and few human cancer cell lines (Yáñez and Porter, 1998; Adachi, Nishijima and Shibahara, 2008; Bouabe and Okkenhaug, 2013).

The efficiency of gene targeting was successfully increased in the following years, following the utilisation of endonucleases from the budding yeast *Saccharomyces cereviciae* to introduce a double-strand break (DSB) at the target site (Jasin, 1996). Examples include the use of the HO endonuclease (Rudin, Sugarman and Haber, 1989) and the I-SceI meganuclease (Rouet, Smih and Jasin, 1994; Choulika *et al.*, 1995). The resulting DSBs promoted the occurrence of HR on the target site and thus increased the efficiency of gene targeting by several orders of magnitude. Although this work demonstrated the key underlying principle of genome editing, each of the above endonucleases has a specific recognition site that determines the site of DNA cleavage. Therefore, this approach could not readily be applied to other loci where the particular recognition sequence of the endonuclease is absent.

1.1.2 Genome editing with programmable nucleases

In order to overcome the limitation of natural yeast endonucleases for their DNA recognition sites, scientists opted for engineering alternative versions of endonucleases that could be targeted to selected genomic locations.

The Zinc Finger Nuclease (ZFN) was the initial system of 'designer nucleases' that was developed (Kim, Cha and Chandrasegaran, 1996; Bibikova *et al.*, 2001, 2003; Pabo, Peisach and Grant, 2001; Porteus and Baltimore, 2003). These 'designer nucleases' contain the catalytic domain of the FokI restriction endonuclease, which is responsible for generating DSBs, and a varying set of three to six zinc finger domains that can be customised to direct the ZFNs to any desirable site in the target genome (Fu and Voytas, 2012). Each zinc finger consists of a Cys2 His2-type zinc finger domain of a transcription factor and typically recognises three nucleotides in the genome (Durai *et al.*, 2005). **Figure 1.1.A** illustrates the ZFN genome editing strategy. The FokI fusion domain is capable of cleaving only one of the strands of the DNA double helix, therefore a pair of ZFNs positioned in proximal sites are required to generate a DSB at the intended target site (Urnov *et al.*, 2005).

Despite overcoming the need to rely on endogenous recognition sites for natural endonucleases, ZFNs pose additional limitations. Customised ZFNs are not cost-effective because they are produced commercially due to the demanding design process (Hermann *et al.*, 2014). In addition, the final constructs may not always be functional (Ramirez *et al.*, 2008) and are prone to associating with DNA sequences that resemble, but are not identical to the intended target site (Gupta *et al.*, 2011).

A second system of 'designer nucleases' was later developed as a more cost-effective alternative to ZFNs. Transcription activator-like effectors (TALEs) are proteins that naturally exist within bacterial plant pathogens of the genus *Xanthomonas.* In order to increase their chances of survival,

Xanthomonas inject TALEs into their host plant cells, where the TALE proteins regulate gene expression to create a favourable environment for the bacterial pathogens (Boch and Bonas, 2010; Bogdanove, Schornack and Lahaye, 2010).

Similar to ZFNs, TALEs have been fused to the catalytic domain of the FokI endonuclease, to generate TALENs, which are programmable nucleases capable of performing targeted genome editing in a genomic site of choice (Morbitzer *et al.*, 2010; Li, Huang, Jiang, *et al.*, 2011; Li, Huang, Zhao, *et al.*, 2011; Mahfouz *et al.*, 2011; Miller *et al.*, 2011). As shown in **Figure 1.1.B**, a pair of TALENs is required to generate a DSB and trigger DNA repair via HR. TALENs can be produced within research laboratories by individual scientists with the aid of several available cost-effective methods for design and assembly (Cermak *et al.*, 2011; Zhang *et al.*, 2011; Garg *et al.*, 2012; Sanjana *et al.*, 2012).

Custom programmable nucleases have enabled scientists to create a plethora of specific edits in multiple model systems (Urnov *et al.*, 2010; Joung and Sander, 2013). Despite the inefficiency of traditional gene targeting in human embryonic stem cells (hESCs), ZFNs and TALENs have been efficiently used to engineer a variety of human pluripotent cell lines (Hockemeyer *et al.*, 2009, 2011). 'Designer nucleases' have evidently paved the way for programmable genome editing and established the potential for future uses of genome engineering in medicine and biotechnology. The design and construction of these constructs remains laborious, expensive and time-consuming (Larson *et al.*, 2013), which imposes limits on the scalability of their production. However, these issues were overcome after the development of a novel 'RNA-guided' system, derived from a naturally occurring mechanism of immunity in prokaryotes, which truly revolutionised all prior strategies for generating precise edits to genomic locations (Menke, 2013; Hsu, Lander and Zhang, 2014).

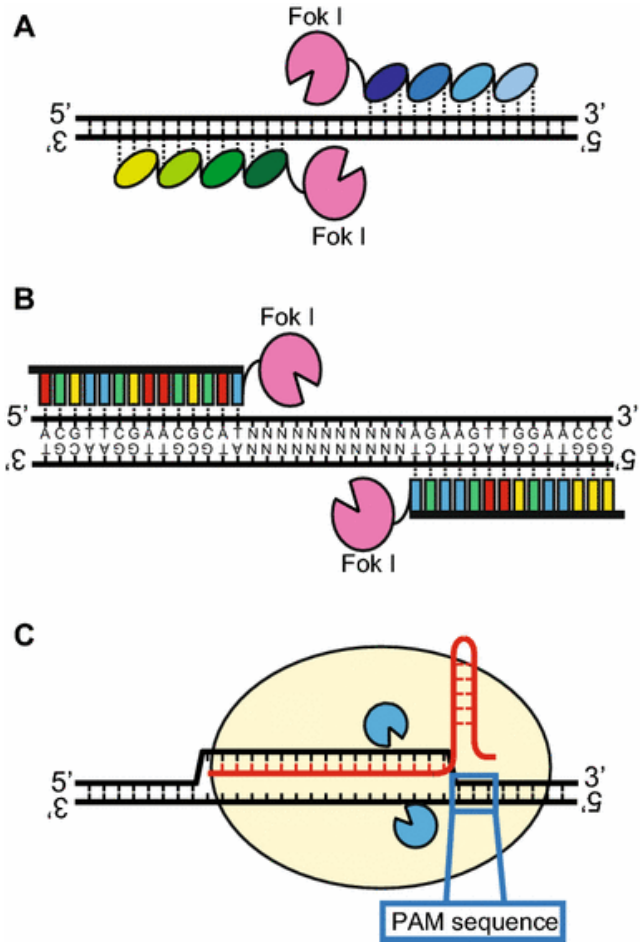


Figure 1.1 Schematic of DSB formation by different programmable nucleases.

A: Paired ZFNs in action. Ellipses represent the four distinguishable zinc finger domains (as an example) while associating with a nucleotide triplet. The catalytic domain of the FokI restriction endonuclease is shown in pink. **B:** Similar to ZFNs, a pair of TALENs is required to generate a DSB, and the fused FokI domain is shown in pink. DNA binding specificity is determined by a core repetitive domain containing repeat-variable di-residues (RVDs). Each RVD associates with a distinct nucleotide of the target site. Amino acids NI (red), HD (green), NG (blue) and NN (yellow) target adenine, cytosine, thymine and guanine, respectively. **C:** Diagram of the target site recognition of CRISPR/Cas9. Target specificity is determined by the sgRNA molecule (shown in red) and the presence of a PAM sequence (highlighted in blue). A DSB is formed through the combined action of the two nuclease domains of Cas9, as shown in blue. **Adapted from Tokunaga *et al*., 2016.**

1.1.3 Genome editing with CRISPR/Cas9

CRISPR/Cas9 stands for Clustered Regulatory Interspaced Short Palindromic Repeats/ CRISPR-associated protein 9 and comprises an RNA-guided endonuclease (Cas9) and a single-guide RNA (sgRNA) molecule (Mali *et al.*, 2013; Cong *et al.*, 2013; Jinek *et al.*, 2013).

The origin of the most popular and widely-used genome editing tool lies in microbial adaptive immunity mechanisms that have been discovered in bacteria and archaea (Makarova *et al.*, 2011). The natural function of CRISPR/Cas systems is to protect the host microorganism from predatory viruses and invading nucleic acids of foreign origin (Horvath and Barrangou, 2010; Bhaya, Davison and Barrangou, 2011). The CRISPR/Cas systems that have been characterised in detail until now have been classified into three distinct categories (Makarova *et al.*, 2015).

The common characteristics of all CRISPR/Cas defence mechanisms are the genes that encode RNA-guided Cas nucleases, non-coding RNAs and a characteristic array of repetitive elements. Short fragments of the invading DNA, or spacers, are integrated into this unique array in an interspaced manner. Following transcription of the entire array and generation of small guide RNA molecules, the unique short spacer sequences allow the host microorganism to cope with a consecutive round of infection from the same genetic material. In general, the guide RNA molecule leads the Cas nuclease to the corresponding complementary re-invading foreign nucleic acid, which is consecutively cleaved by the Cas nuclease (Garneau *et al.*, 2010). In addition, cleavage of an invading DNA molecule requires the presence of a short characteristic motif that helps the microorganism determine its foreign origin (Mojica *et al.*, 2009). This is called the ‘protospacer-adjacent motif’ or PAM and is equally important in the function of the prevalent CRISPR/Cas9 genome editing tool.

The most commonly used version of CRISPR/Cas was developed from the Type II immunity system from *Streptococcus pyogenes* and consists of the Cas9 endonuclease, the crRNA array of spacers and the trans-activating crRNA (tracrRNA). The tracrRNA molecules hybridise with the repetitive portions of the pre-crRNA. RNaseIII is responsible for separating the crRNA-tracrRNA duplexes into distinct molecules. Every duplex then binds Cas9 and determines the specificity of the intended DSB site and positions Cas9 there through a 20 nucleotide-long protospacer sequence. The presence of an adjacent 5′ NGG PAM allows base pairing between the protospacer of the crRNA and its complementary target site. If sufficient complementarity is present in this region, Cas9 undergoes a conformational transition that allows the HNH and RuvC nuclease domains of Cas9 to cleave the complementary and non-complementary strand, respectively, and generate a targeted DSB between positions 3 and 4 of the protospacer sequence (Jinek *et al.*, 2012; Sternberg *et al.*, 2014; Jiang *et al.*, 2016). A schematic is provided in **Figure 1.1.C**. In addition, the domain organisation and crystal structure of Cas9 from *Streptococcus pyogenes* are shown in **Figure 1.2.A** and **Figure 1.2.C&D**, respectively.

Repurposing of the *Streptococcus pyogenes* CRISPR/Cas9 system into the most potent genome editing tool was taken forward by the fusion of the crRNA and tracrRNA components to generate a chimeric single guide RNA (sgRNA) molecule, shown in **Figure 1.2.B**. This reduced the number of separate components required for editing to two, and led to the establishment of a novel programmable nuclease which was immediately favoured for generating a plethora of edits in mammalian cell lines (Cho *et al.*, 2013; Cong *et al.*, 2013; Jinek *et al.*, 2013; Mali *et al.*, 2013).

The main advantage of CRISPR/Cas9 is that the target site specificity is determined solely from the protospacer portion of the sgRNA. The protospacer sequence can be easily modified by combining oligonucleotide synthesis with standard molecular cloning, thus rendering Cas9 reprograming easy and affordable for many scientists (Ran, Hsu, Wright, *et al.*, 2013). The production cost of CRISPR/Cas9 was greatly reduced compared to ZFNs and TALENs.

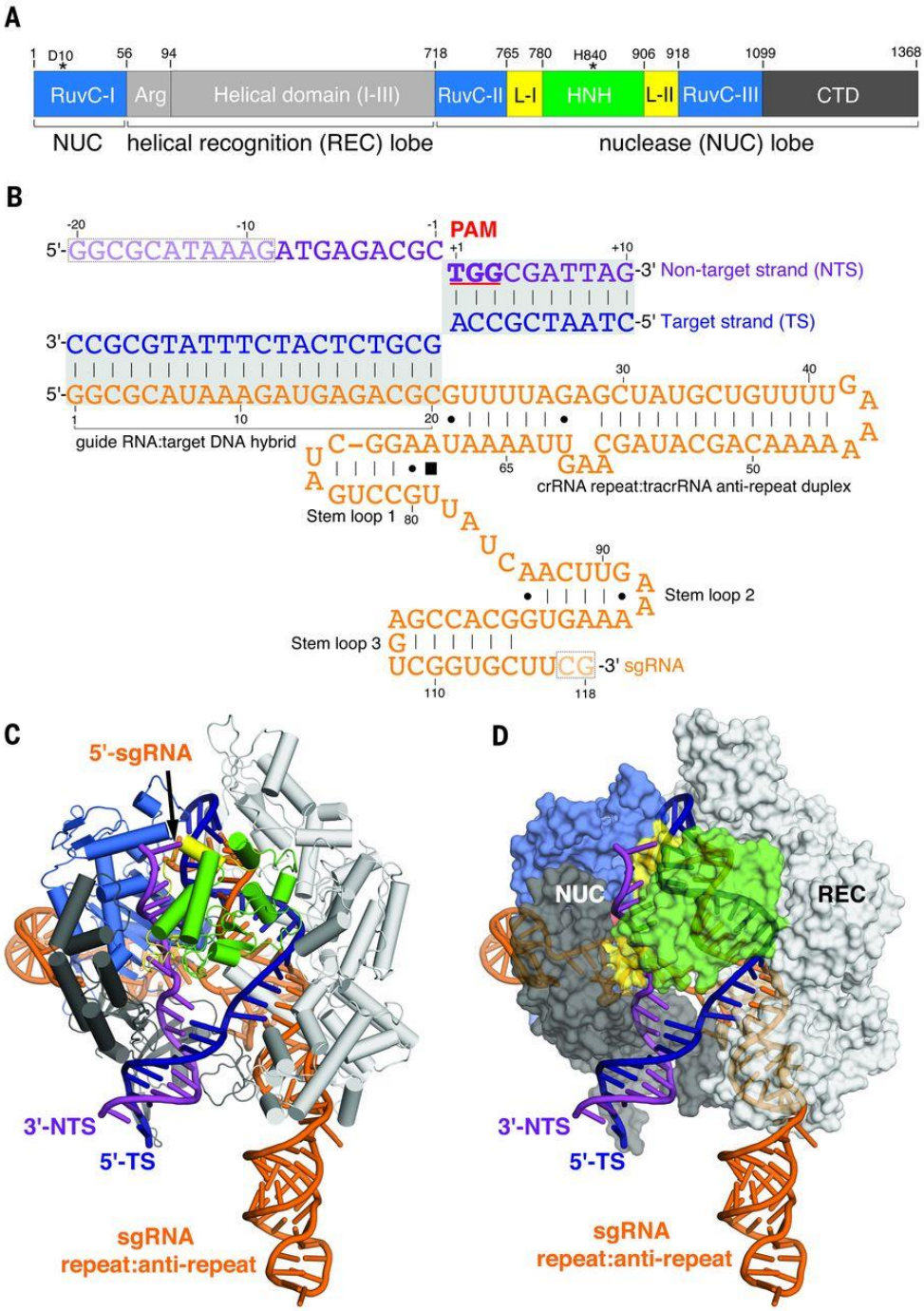


Figure 1.2 Crystal structure of the *S. pyogenes* Cas9-sgRNA complex in action.

A: Linear representation of the domains of Cas9 from S. pyogenes. **B:** Diagram of the chimeric sgRNA molecule when it associates with its complementary target DNA. **C&D:** The Cas9-sgRNA complex upon association with the target site. The crystal structure is shown in ribbon and surface representation, respectively. The Cas9 domains and sgRNA are colour-coded as for A and B above. **Adapted from Jiang *et al*., 2016.**

Moreover, co-expression of multiple sgRNAs with the Cas9 nuclease allowed multiple genomic locations to be targeted simultaneously in the same cell. Pools consisting of thousands of sgRNAs can be introduced simultaneously into cell populations through lentiviral delivery (Koike-Yusa *et al.*, 2014). This created ample opportunities for high-throughput screens and additional modifications such as rearrangements and large-scale chromosomal deletions (Cong *et al.*, 2013; Choi and Meyerson, 2014; Li *et al.*, 2014; Shalem *et al.*, 2014; Torres *et al.*, 2014; Wang *et al.*, 2014; Chen *et al.*, 2015; Kraft *et al.*, 2015; Doench *et al.*, 2016).

One key disadvantage of CRISPR/Cas9 is the tolerance of mismatches during pairing of the protospacer of the sgRNA and genomic locations that highly resemble its intended target site. Mismatch tolerance usually leads to a high frequency of off-target hits and makes CRISPR/Cas9 less specific than other versions of programmable nucleases, most notably TALENs (Li *et al.*, 2014). Nonetheless, a series of adaptations have been made to increase the specificity and efficiency and scope of CRISPR/Cas9 as a genome editing tool (Cebrian-Serrano and Davies, 2017; Komor, Badran and Liu, 2017). Cas9 nickases (Cas9n) have been generated to be used in a pairwise manner that resembles ZFNs and TALENs, thus the size and fidelity of the recognition site was doubled (Ran, Hsu, Lin, *et al.*, 2013; Shen *et al.*, 2014).

Over the years, multiple algorithms and tools have emerged to predict the likelihood of off-target site recognition for individual sgRNAs (Tsai *et al.*, 2015, 2017; Kim *et al.*, 2016; Cameron *et al.*, 2017; Kim and Kim, 2018; Wienert *et al.*, 2019). Scientists can thus minimize the chances of getting off-target hits during the experimental design stage. Several cell-free, but also cell-based, off-target detection methods have been developed to predict putative off-target hits for CRISPR/Cas9. In general, cell-free methods produce large lists consisting of hundreds of possible off-target sites for every individual sgRNA that is examined (Kim *et al.*, 2016; Cameron *et al.*, 2017; Tsai *et al.*, 2017). Subsequent experimental validation of these predicted off-target sites frequently results into the realisation that many of them are false positives. By

contrast, cell-based off-target detection methods produce smaller lists of predicted off-target sites (Tsai *et al.*, 2015).

A couple of more recent CRISPR/Cas9 off-target site prediction methods have highlighted the importance of integrating information about chromatin organisation relative to the target site. DIG-seq is a cell-free method, where chromatinised genomic DNA is incubated with purified Cas9 protein and sgRNAs *in vitro* (Kim and Kim, 2018). This method is a more recent development of Digenome-seq, a cell-free method that also involved purified genomic DNA without histones or other binding protein factors. Similar to cell-based methods, DIG-seq produces a smaller list of putative off-target sites for the tested sgRNAs, but still fails to capture what happens in a real biological setting, where chromatin is dynamic.

The latest development amongst off-target detection methods is DISCOVER-seq, which is cell-based but also applicable to mice. This astonishing method allows detection of Cas9-induced DSB through MRE11 ChIP following exposure to CRISPR/Cas9. DISCOVER-seq is widely applicable to mammalian species used heavily in biological research, due to the conservation of MRE11, and allows reliable DSB detection with minimal false positives (Wienert *et al.*, 2019). This is true because MRE11 is one of the first factors that associate with the ends of Cas9-induced DSBs, as mentioned later in section **1.2.1** in this chapter, during DSB repair through HR, alt-EJ and SSA (Haince *et al.*, 2008; Williams *et al.*, 2008; Williams, Lees-Miller and Tainer, 2010). MRE11 association may not be relatable with repair of Cas9-induced DSBs through c-NHEJ but is yet the best method for reporting on DSB recognition instead of the endpoint of CRISPR/Cas9 mutagenesis process.

This dissertation will be focusing on genome editing with the traditional Cas9 endonuclease from *Streptococcus pyogenes*. In the context of genome editing in mammalian cells, Cas9 generates targeted DSBs that are sealed by inherent DNA repair mechanisms.

1.2 DNA double-strand break repair, CRISPR/Cas9 and chromatin

1.2.1 DSB repair pathways in eukaryotes in the context of genome editing

After the DSB has been generated by a genome editing nuclease, changes to the DNA sequence occur during their repair by the native DNA repair machinery of the cell. The two main pathways of DNA DSB repair found in eukaryotes are homologous recombination (HR) and non-homologous end joining (NHEJ). The term Homology Directed Repair (HDR) is used to describe the process of DSB repair with the aid of an exogenous template donor during genome editing (Doudna and Charpentier, 2014), and will be discussed in more detail later in this section. While HR always leads to complete restoration of the DNA sequence in the immediate vicinity of the DSB, NHEJ is the mechanism favoured by multicelluar eukaryotes for DSB repair (Lieber *et al.*, 2003; San Filippo, Sung and Klein, 2008).

Fidelity during DNA DSB repair is of paramount importance. This is why the principle of the error-free HR mechanism is conserved from prokaryotes to mammals (Mladenov *et al.*, 2016). After a DSB is identified, the MRN complex, cosisting of the Mre11, Rad50 and Nbs1 factors, is directed to the free DNA ends (Williams *et al.*, 2008; Williams, Lees-Miller and Tainer, 2010). This event is called DSB sensing and triggers recruitment of the CtIP nuclease, which in turn begins processing of the DNA ends (Eid *et al.*, 2010; Makharashvili and Paull, 2015). The 5´ to 3´ exonuclease Exo1 and Bloom helicase (BLM) are also recruited to unwind the DNA, complete end processing and lead to the formation of a 3´ end overhang on both ends (Bischof *et al.*, 2001; Zakharyevich *et al.*, 2010). The replication protein A (RPA) subsequently coats both 3´ overhangs and attracts the PALB2-BRCA1-BRCA2 complex (Moynahan, Pierce and Jasin, 2001; Xia *et al.*, 2006; Prakash *et al.*, 2015). As a result, BRCA2 mediates the exchange of RPA with Rad51, a factor that is

responsible for invasion of the 3´ overhang into the homologous region on the sister chromatid (Esashi *et al.*, 2005). Several paralogs of Rad51 come on site to aid with resolution of the Holiday junction that is formed as part of HR (Masson *et al.*, 2001). The most notable of them is Rad54, which is implicated in branch migration and Holiday juntion resolution (Bugreev, Hanaoka and Mazin, 2007; Mazina and Mazin, 2008; Mazin *et al.*, 2010). Resolvases, such GEN1 found in human, complete the process of Holiday junction resolution, leading to error-free repair of the broken sister chromatid (Ip *et al.*, 2008). Due to its ability to fully restore the integrity of a broken DNA molecule, HR also is involved in other cellular processes where DSBs are induced, such as meiotic recombination (Inagaki, Schoenmakers and Baarends, 2010; Schatz and Swanson, 2011).

HR fundamentally relies on the presence of a homologous sequence, typically the sister chromatid, in order to repair a DSB occuring on a chromosome. HR during G1 with a homolog rather than a sister chromatid leads to loss of heterozygosity, with a strong potential for downstream complications such as development of cancer (Stark and Jasin, 2003). Therefore, HR is temporally restricted to the S and G2 phases of the cell cycle through a mechanism that involves ubiquitination of PALB2 to abrogate its interaction with BRCA1 during G1 (Orthwein *et al.*, 2015). Another property of this high-fidelity mechanism is that HR is time consuming because it involves numerous steps of processing and resolution (Mladenov *et al.*, 2013).

On the contrary, c-NHEJ has a very short half-time of 10 to 30 minutes and can occur throughout the cell cycle, making it the more frequently utilised mechanism to repair DSBs in most animal cells (Iliakis *et al.*, 2004). During classical NHEJ (c-NHEJ) the ends of a DSB are initially bound by the Ku70/80 complex, which can excise damaged nucleotides and recruit the DNA-dependent protein kinase (DNA-PKcs) (Downs and Jackson, 2004; Roberts *et al.*, 2010; Strande *et al.*, 2012). After DNA-PKcs associates with the Ku70/80 complex it autophosphorylates and marks the beginning of the repair process (Chan *et al.*, 2002). Autophosphorytaltion of DNA-PKcs leads to the recruitment of the Artemis 5´ to 3´ exonuclease, which also trims away

damaged nucleotides (Moshous *et al.*, 2001; Goodarzi *et al.*, 2006). In addition, the polμ and polλ polymerases are also recruited by autophosphorylated DNA-PKcs to fill in the small gaps that are created during processing by Artemis (Daley *et al.*, 2005; Paull, 2005). These processing steps ensure that the DNA ends become compatible for ligation, which occurs via the LigaseIV DNA ligase after the ends are stably positioned in close proximity by the XRCC4 and XLF factors (Andres *et al.*, 2012; Roy *et al.*, 2015). These factors do not have any enzymatic activity but play a key role towards bridging the ligatable DNA ends. PAXX, a paralog of XRCC4, and XLF are also involved in bridging the processed ligatable ends by directly associating with the Ku7/80 complex which is still found on either end (Ochi *et al.*, 2015). This mechanism can lead to error-free DSB repair, but c-NHEJ is also associated with small single-nucleotide insertions or deletions (InDels) that arise as a consequence of this type of DNA end processing (Mladenov *et al.*, 2016).

Over the years, additional mechanisms that resemble c-NHEJ have been discovered to repair DSBs in the event of c-NHEJ inhibition. Alternative End Joining (alt-EJ) relies on the recruitment of PARP1 instead of the Ku70/80 complex to initiate DSB repair (Wang *et al.*, 2006). Histone H1 has been reported to enhance the activity of PARP1 during alt-EJ (Rosidi *et al.*, 2008). End processing occurs with the aid of the MRN complex and CtIP nuclease, which are primarily part of the HR pathway (Haince *et al.*, 2008). Polθ is subsequenctly recuited to prevent association of Rad51 and fill in the single-stranded DNA overhangs generated by MRN-CtIP (Ceccaldi, Rondinelli and D'Andrea, 2016). During alt-EJ ligation of the processed DNA ends is achieved by the combined action of the DNA LigaseI and DNA LigaseIII and their cofactor XRCC1 (Audebert, Salles and Calsou, 2004; Arakawa and Iliakis, 2015).

Extended resection facilitated by MRN-CtIP during alt-EJ may reveal short sequences of homology ranging from 2-25 nucleotides. Alignment of such microhomologies frequently leads to deletion of intermediate sequences through a process that is refered to as microhomology-mediated end joining (MMEJ), which is highly mutagenic and has even been associated with large

chromosomal translocations (McVey and Lee, 2008; Villarreal *et al.*, 2012). Indeed, the presence of such large deletions that are reminiscent of MMEJ has also been observed after genome editing with CRISPR/Cas9 (Kosicki, Tomberg and Bradley, 2018). Evidence from a recent study comparing the repair kinetics of Cas9-induced and irradiation-induced DSBs in the absence of an externally provided repair donor concluded that repair of Cas9-induced DSBs requires more time on average and may last up to 15 hours (Brinkman *et al.*, 2018). Furthermore, chemical inhibition of DNA-PKcs resulted primarily in larger deletions of up to 10 nucleotides, leading the authors to conclude that MMEJ works as a backup mechanism of c-NHEJ during genome editing with CRISPR/Cas9 (Brinkman *et al.*, 2018). A diagram illustrating the spectrum of InDels that can be generated during error-prone NHEJ pathways is provided in **Figure 1.3**.

Nevertheless, genome editing is typically used to introduce precise sequence changes to a genomic location of choice through HDR. This requires co-delivery of a single-stranded or double-stranded exogenous template bearing homology arms, as shown in **Figure 1.3**. Emerging evidence suggests that the nature of the repair donor determines the pathway of DSB repair, since single-stranded templates have been reported to stimulate DSB repair in the absence of the BRCA2 mediator of the traditional HR pathway (Kass *et al.*, 2016; Bothmer *et al.*, 2017).

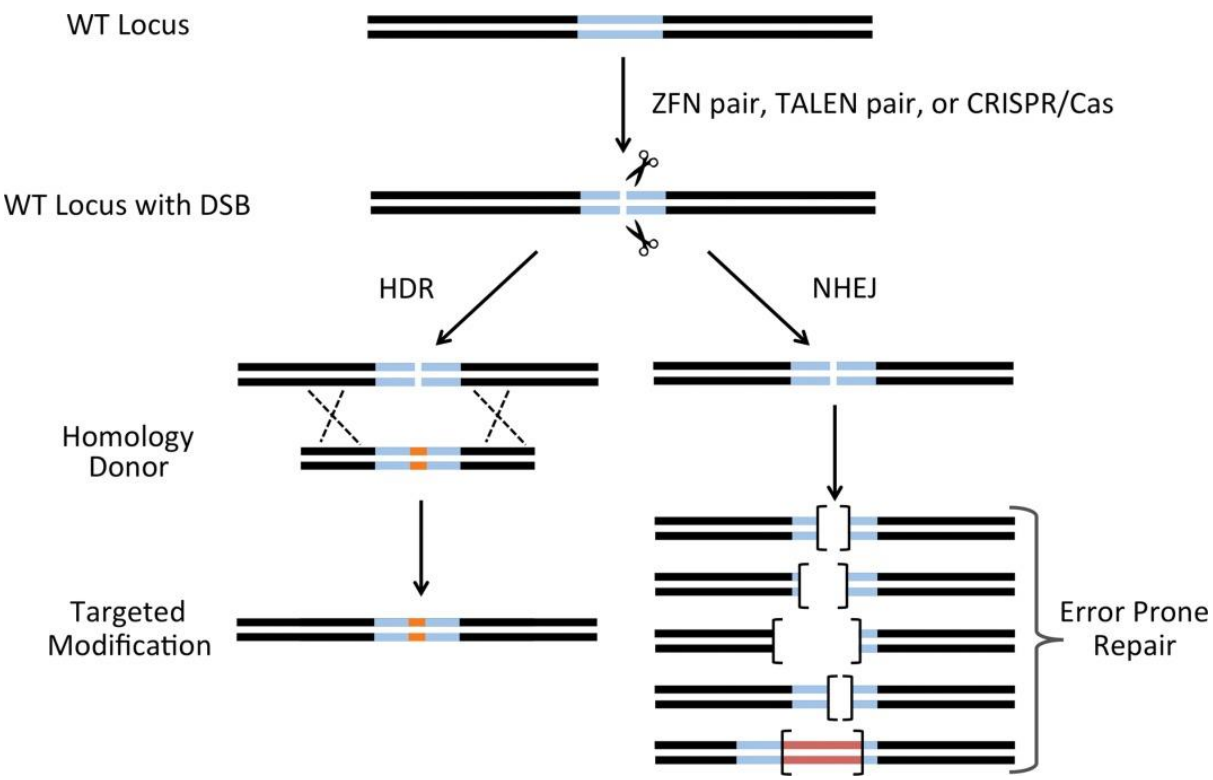


Figure 1.3 Double-strand break repair pathways during genome editing with programmable nucleases.

ZFNs, TALENs and CRISPR/Cas9 generate DSBs that are repaired either by the error-prone NHEJ mechanism or by HDR when a homologous donor is supplied. Utilisation of HDR leads to the introduction of specific desirable modifications (shown in orange) on the target site (shown in light blue). Alternatively, c-NHEJ and MMEJ may lead InDel formation on the target site. Error-free repair by c-NHEJ is also highly possible. InDel size (indicated by brackets) is determined by the error-prone pathway that is utilised to seal the DSB. **Adapted from Menke, 2013.**

While repair through double-stranded donors appears to utilise common factors in common with classical HR such as Rad51 (Bothmer *et al.*, 2017), there is a general consensus that the use of single-stranded donors during genome editing leads to DSB repair through a mechanism that resembles synthesis-dependent strand annealing (SDSA). Gene conversion tract analysis revealed that Cas9 tracts are unidirectional, suggesting that one of the DNA ends associates, synthesises from and then dissociates with the single-stranded donor, in a manner that resembles SDSA and does not require on incorporation of the actual donor molecule (LaRocque and Jasin, 2010; Kan *et al.*, 2017). However, HDR with single-stranded donors during genome editing is Rad52-dependent and therefore associated with single-strand annealing (SSA), an error-prone homology-driven DSB repair pathway that was originally discovered in unicellular eukaryotes and prokaryotes (Fishman-Lobell, Rudin and Haber, 1992; Malkova *et al.*, 2001; Cubbon, Ivancic-Bace and Bolt, 2018; Gallagher and Haber, 2018). During SSA Rad52 coats and bridges the processed DNA ends. This rough process results in flaps that are subsequently nicked by the ERCC1/XPF nuclease. Finally, the DNA LigaseI seals the nicks but leads to variable-sized deletions (Symington, 2002; Ceccaldi, Rondinelli and D'Andrea, 2016).

Additional evidence has recently emerged suggesting that HDR with single-stranded repair donors is also linked with the Fanconi Anemia (FA) pathway. Depletion of FA factors during a targeted CRISPRi screen increased the frequency of NHEJ utilisation. Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) revealed that the FANCD2 FA factor collocalises with Cas9 at DSBs during genome editing with CRISPR/Cas9 (Richardson *et al.*, 2018). As characterisation of the HDR mechanism used during Cas9-induced DSB repair with single-stranded repair templates remains incomplete, it is also worth considering that it occurs in an environment where the DNA is not freely accessible.

1.2.2 Parameters that affect DSB repair within eukaryotic cells

Unlike prokaryotes, eukaryotes sequester their genome inside a nuclear membrane. Certain regions within the eukaryotic nucleus appear very compacted and dense and are historically referred to as heterochromatin (Brown, 1966; Straub, 2003). Heterochromatin is described in more detail in section **1.3.1** of this chapter. It has recently been shown that the position where a DSB occurs within the nucleus determines whether HR or NHEJ will be utilised for repair. If a DSB occurs within a region close to the nuclear membrane it is more likely to be repaired by NHEJ. This is because the regions lining the interior of the nuclear membrane are A/T-rich, non-transcribed during interphase and heterochromatic (Fawcett, 1966; Meuleman *et al.*, 2013; van Steensel and Belmont, 2017). On the contrary, DSBs occurring in the interior of the nucleus or in close proximity to the nuclear pore can be repaired either by HR or NHEJ (Lemaître *et al.*, 2014).

Further evidence from two individual groups suggests that DNA ends from DSBs arising in DAPI-dense nuclear compartments migrate towards the periphery of these, where they are repaired by HR (Chiolo *et al.*, 2011; Jakob *et al.*, 2011). A more recent study also verified the above by generating targeted DSBs within chromocenters with CRISPR/Cas9. Chemical synchronisation of mouse NIH 3T3 cells in G2 allowed a direct comparison between the chances of NHEJ and HR occurrence. Monitoring of DNA end localisation through structured illumination microscopy (SIM) revealed that DSBs generated within DAPI-dense regions relocate towards the periphery of these regions, where they co-localise with RAD51. DNA end migration requires processing to have occurred, since depletion of MRE11 and CtIP results in stationary DSBs as assessed by immunofluorescence (IF) monitoring of the DNA end marker γH2Ax (Tsouroula *et al.*, 2016). The same study showed that DSBs escaping relocation are repaired by SSA within heterochromatin in a RAD52-dependent manner in the absence of BRCA2.

A link between transcription and DSB repair has also emerged since it was observed that actively transcribed genomic regions are preferably repaired by HR. An engineered U2OS cell line, where targeted DSBs are generated in an inducible manner via the AsiSI restriction endonuclease, was used to study the relationship between DSBs, epigenetic modifications and transcription (Aymard *et al.*, 2014). Chemical inhibition of transcription resulted in decreased accumulation of RAD51 near generated DSBs. The lens epithelium–derived growth factor (LEDGF) normally interacts with H3K36me3, a histone post-translational modification deposited within actively transcribed gene bodies, and recruits CtIP (Daugaard *et al.*, 2012). Depletion of LEDGF also led to a dramatic decrease of RAD51 accumulation on DSB sites located in actively transcribed genes. RNAi-mediated depletion of SETD2, the methyltransferase responsible for H3K36me3 deposition, resulted in the same reduction in RAD51 accumulation at the same DSB sites. Therefore, it was concluded that DSBs occurring within transcriptionally active genes are preferentially repaired by HR (Aymard *et al.*, 2014).

The idea that other factors, such as eukaryotic chromatin compaction, affect DNA DSB repair pathway choice is becoming increasingly popular (Clouaire and Legube, 2015). The term chromatin compaction is used frequently throughout every chapter in this thesis. This term is used to describe the different features of chromatin in eukaryotes, such as epigenetic modifications and accessibility to DNase enzymes, as well as transcription. In general, I use the term ‘compacted’ chromatin to describe a region that is transcriptionally silent, bears repressive histone modifications and methylated CG dinucleotides. A description for all these eukaryotic chromatin features is provided in the following section. Since it is unclear how Cas9-induced DSBs are repaired with the aid of a single-stranded donor molecule, it would be interesting to investigate if pre-existing chromatin compaction affects the choice between NHEJ and HDR during genome editing with CRISPR/Cas9.

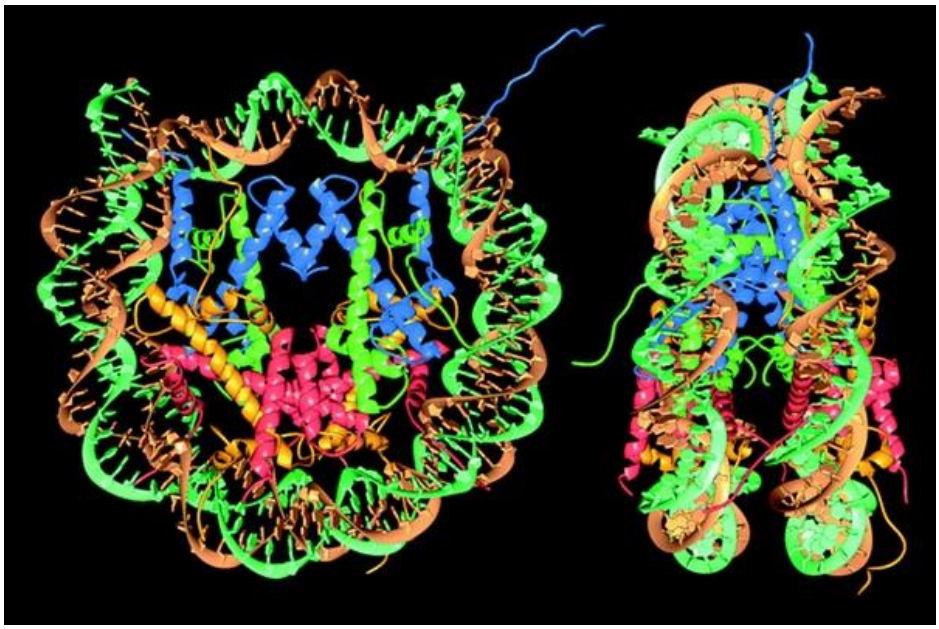
1.3 Chromatin characteristics and genome editing

1.3.1 Chromatin features in eukaryotes

The genomes of higher eukaryotes are compacted into chromatin in order to fit inside the nucleus (Woodcock and Ghosh, 2010). The 'fundamental subunit' of this higher order structure is the nucleosome, which consists of 147 base pairs that are wrapped around a core histone protein octamer, as shown in **Figure 1.4.A** (Kornberg, 1974; Kornberg and Thonmas, 1974; Luger *et al.*, 1997). Inter-nucleosomal interactions result in the formation of chromatin fibres with higher order structures, which in turn are compacted further up to mitotic chromosomes, as shown in **Figure 1.4.B** (Woodcock and Dimitrov, 2001; Baylin and Schuebel, 2007). It has been demonstrated that nucleosomes inhibit transcription initiation but may be displaced by RNA Polymerase II during transcriptional elongation *in vitro* (Knezetic and Luse, 1986; Lorch, LaPointe and Kornberg, 1987).

Chromatin can bear additional layers of chemical modifications. Methyl groups can be added to the fifth carbon atom of cytosine residues within primarily CG dinucleotides in eukaryotes (Zemach *et al.*, 2010). This modification is termed DNA methylation and it is the only know modification to be directly deposited into the genome in mammals (Robertson, 2005). Mammals poses two main DNA methyltransferases that deposit the mark *de novo* in any DNA sequence context, DNMT3A and DNMT3B (Okano, Xie and Li, 1998; Okano *et al.*, 1999; Greenberg and Bourc'his, 2019). DNMT3L is a third de novo DNA methyltransferase (Bourc'his *et al.*, 2001; Ooi *et al.*, 2007), which is only catalytically active in the germline and will be mentioned later in **section 1.4**. During DNA replication the DNA methylation mark is not passed on to the daughter DNA strand, resulting in hemimethylated CG sites. These hemimethylated CG sites are bound by UHFR1, or Np95 in mouse, which in

A



B

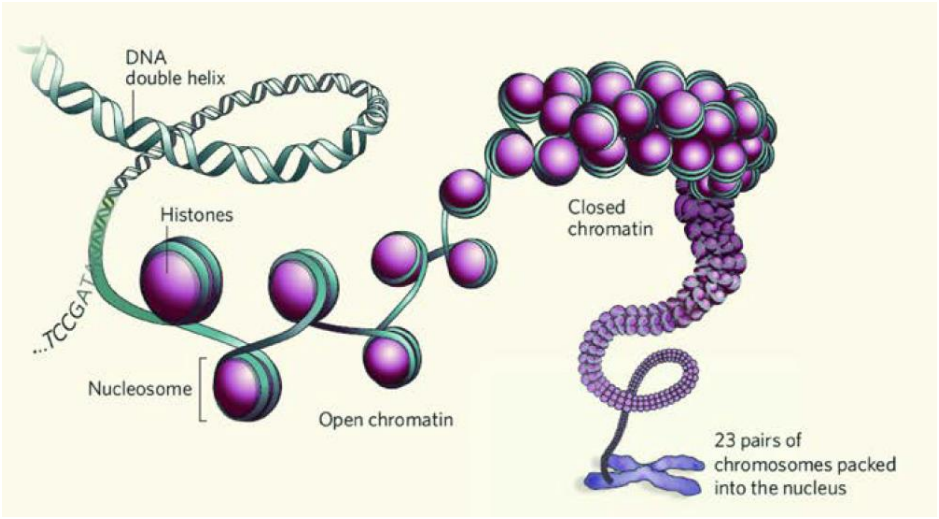


Figure 1.4 Chromatin structure representations. A: The crystal structure of a nucleosome, where the DNA double helix surrounds the histone octamer. Histones H2A, H2B, H3 and H4 are represented in orange, red, green and blue, respectively. **Adapted from Luger *et al.*, 1997. B:** Representation of the different levels of chromatin organisation. From left to right there is the naked DNA double helix, the 'beads on a string' nucleosome fibre, which is then condensed into a larger fibre, through to a highly compacted mitotic chromosome. **Adapted from** mitotic chromosomes. **Adapted from Baylin and Schuebel, 2007.**

turn attracts the maintenance DNA methyltransferase DNMT1 to deposit methyl groups to the relevant cytosine residues of the newly synthesised DNA strand (Bostick *et al.*, 2007; Sharif *et al.*, 2007; Ishiyama *et al.*, 2017). Finally, DNA methylation is removed by TET enzymes, which facilitate oxidation of 5-methylcytosine residues (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Tahiliani *et al.*, 2009; Ito *et al.*, 2010, 2011; He *et al.*, 2011). All oxidised forms of 5mC lead to loss of DNA methylation maintenance after DNA replication (Otani *et al.*, 2013). Furthermore, 5fC and 5caC can also by removed through the base excision pathway since both are substrates for the thymine DNA glycosylase enzyme (Weber *et al.*, 2016).

In addition, histone proteins within nucleosomes can be post translationally modified to affect chromatin structure and function. Histone tails protrude through the nucleosome and act as substrates for a repertoire of modifications (Mersfelder and Parthun, 2006). The most notable ones that will be discussed here are histone acetylation and histone methylation.

Histone acetylation causes chromatin relaxation by neutralising the positive charge of lysine residues, where an acetyl group is deposited. This modification is associated with transcriptionally active genomic regions (Allfrey, Faulkner and Mirsky, 1964; Hebbes, Thorne and Crane-Robinson, 1988; Wang *et al.*, 2009). Chemical inhibition of histone deacetylase enzymes has been shown to lead to an elevated accumulation of acetylated histones, resulting in chromatin decondensation at a global level (Tóth *et al.*, 2004).

Histone methylation is associated both with transcriptionally active and silent regions of the genome, depending on the residue that this small neutral modification is deposited. Histone tails can be methylated on arginine or lysine residues. More specifically, lysine residues may receive up to three distinct methyl groups. The most outstanding methylated histone signatures that are associated with actively transcribed genes are namely H3K4me3 and H3K36me3. These histone methylation signatures are commonly found among promoters of transcriptionally-active genes or within gene bodies of actively transcribed genes, respectively (Krogan *et al.*, 2003; Wang *et al.*, 2008). On

the contrary, H3K9me1/2/3, H4K20me3 and H3K27me3 are associated with chromatin compaction and are frequently referred to as the repressive histone modifications (Wang *et al.*, 2008). More specifically, the Polycomb repressive complex 2 (PRC2) deposits H3K27me3 leading to transcriptional silencing (Cao *et al.*, 2002). In general, methylation of H3K9 works as a docking site for other proteins that may associate with chromatin, such as the heterochromatin protein 1 (HP1), which leads to the formation of heterochromatin (Bannister *et al.*, 2001; Fischle *et al.*, 2005).

DNA methylation also leads to heterochromatin formation, as the DNA methyltransferase enzymes mentioned earlier interact with other types of enzymes while depositing methyl groups on cytosine residues. For example, DNMT1 interacts with the G9a H3K9 methyltransferase (Estève *et al.*, 2006; Epsztejn-Litman *et al.*, 2008), the HDAC1 histone deacetylase (Fuks *et al.*, 2000) and the LSH chromatin remodeller (Dennis *et al.*, 2001; Myant *et al.*, 2011; Tao *et al.*, 2011). These enzymes can alter the chromatin landscape at methylated DNA sequences, eventually leading to transcriptional repression of underlying genes. Furthermore, the binding of several transcription factors is impaired by DNA methylation (Yin *et al.*, 2017) and a positive correlation between transcriptional repression and the number of methylated CG dinucleotides at gene promoters has also been observed (Weber *et al.*, 2007).

Despite being a small, neutral with respect to charge, modification, DNA methylation works as a substrate for several proteins that can contribute further to heterochromatin formation. In mammals, these include the methyl-CpG-binding protein 2 (MeCP2) and the four identified methyl-CpG-binding domain (MBD) proteins (Meehan *et al.*, 1989; Hendrich and Bird, 1998). These five proteins have been shown to recruit histone deacetylases and contribute to transcriptional silencing of underlying genes (Nan *et al.*, 1998; Ng *et al.*, 1999). Lastly, some zinc finger proteins also associate with methylated cytosines, such as ZFP57 (Quenneville *et al.*, 2011) and Kaiso (Yoon *et al.*, 2003), which lead to DNA methylation maintenance or recruitment of histone deacetylases, respectively. Therefore, all the proteins mentioned above are

not only able to recognise and associate with methylated cytosines, but also mediate heterochromatin formation.

The term heterochromatin was originally created by Emil Heitz to describe chromosomal regions that remained condensed after mitosis (Brown, 1966; Straub, 2003). Further studies demonstrated that heterochromatin is an AT-rich, gene-poor fraction of the genome that is mostly condensed throughout the cell cycle and only replicated during the later stage of S-phase (Brown, 1966; Pardue and Gall, 1970; Comings, 1978; Craig and Bickmore, 1993; Estandarte *et al.*, 2016). Repetitive DNA sequences found on mammalian centromeres and telomeres are examples of sequences that should remain transcriptionally silent and are therefore persistently embedded within constitutive heterochromatin. This type of heterochromatin is very stable and is marked by H3K9me3.

However, if the compaction of a genomic region is dynamic over time, for example during multicellular development, then the term facultative heterochromatin is used to describe the heterochromatic environment where this region is embedded (Dillon and Festenstein, 2002; Trojer and Reinberg, 2007). Examples of such regions include bivalent genes, at which promoter sequences are simultaneously decorated with H3K4me3 and H3K27me3 and play a pivotal role in development and differentiation, as well as the inactive X chromosome in female mammals (Bernstein *et al.*, 2006; Heard and Disteche, 2006).

Chromatin compaction occurs at many levels and is highly dynamic. Despite its natural origins CRISPR/Cas9 is very effective within mammalian cells and seems to perform well in the presence of chromatin. It is though impossible to exclude that CRISPR/Cas9 in somehow influenced by this higher order structure.

1.3.2 CRISPR/Cas9 target search and recognition

Despite its wide usage, the details on the mechanism by which CRISPR/Cas9 searches for and recognises its target are still emerging. A recent study using Förster resonance energy transfer (FRET) demonstrated that CRISPR/Cas9 diffuses in a lateral manner on the DNA double helix and refined previous findings suggesting that the complex diffuses in a three-dimensional manner (Sternberg *et al.*, 2014; Globyte *et al.*, 2019). Once the Cas9-sgRNA complex encounters a PAM site, the Cas9 unwinds the double helix and the sgRNA protospacer hybridises to the complementary target sequence (Sternberg *et al.*, 2014). This results in the formation of an R-loop between the target DNA and sgRNA, which has been proposed to be the rate limiting step for Cas9 cleavage (Gong *et al.*, 2018). During R-loop formation the HNH nuclease domain of Cas9 constantly changes its structural conformation (Dagdas *et al.*, 2017). HNH eventually settles into a permissive state and Cas9 undergoes a conformational change to transition to its active state (Jinek *et al.*, 2014). Following this change, the HNH and RuvC nuclease domains of Cas9 simultaneously cleave the protospacer-bound complementary strand and the non-complementary PAM-containing strand, respectively, to generate a DSB (Jinek *et al.*, 2012). **Figure 1.5** illustrates the Cas9-sgRNA complex in a bound and cleavage-competent conformation with respect to the different levels of chromatin organisation.

However, the above biochemical insights into the CRISPR/Cas9 mechanism of action have been deduced from systems that are not fully representative of what programmable nucleases encounter inside mammalian nuclei. The genomes of higher eukaryotes are compacted into chromatin, which may influence the target search by CRISPR/Cas9, just as it has been shown to influence other nuclear processes such as transcription (Li, Carey and Workman, 2007). In contrast to ZFNs and TALENs, CRISPR/Cas9 did not evolve to recognise chromatinised DNA as a natural substrate.

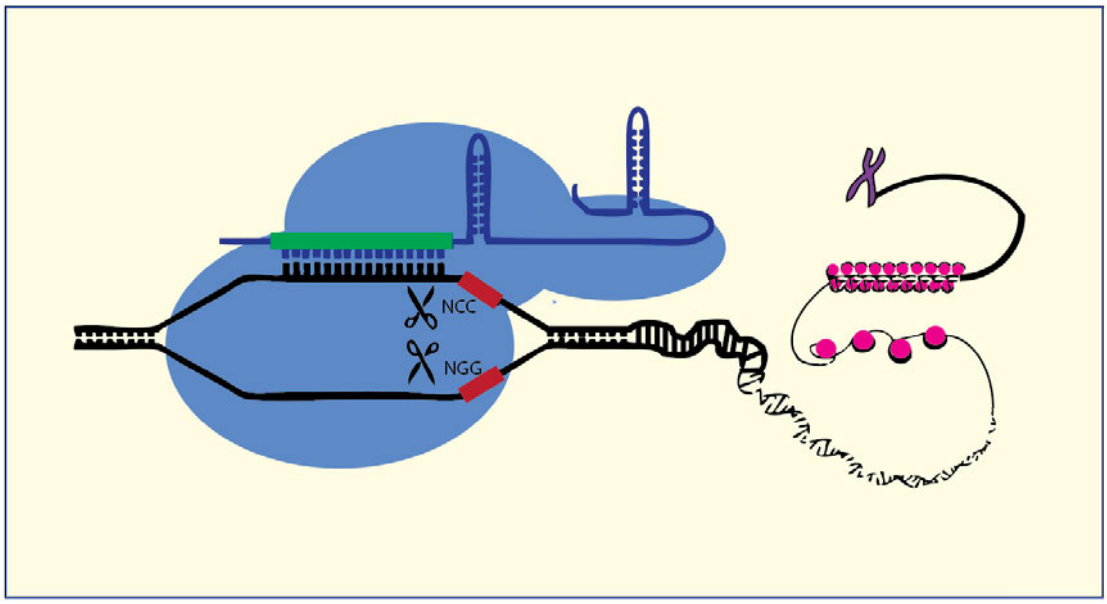


Figure 1.5 CRISPR/Cas9 and the barrier posed by eukaryotic chromatin compaction.

The Cas9-sgRNA complex is shown during the R-loop formation stage, where the protospacer hybridises to the target strand (shown at the top). The pairs of scissors pointing upwards and downwards represent the HNH and RuvC nuclease domains of Cas9, respectively. Cas9 is shown in blue and the sgRNA in dark blue. The sgRNA protospacer and PAM are shown in green and red, respectively. On the right, the different levels of chromatin compaction that eukaryotic genomes undergo are shown. Starting from the DNA double helix at the bottom, to nucleosomes (depicted as pink spheres) forming a chromatin fibre and a fully compacted mitotic chromosome (shown in purple) at the top.

1.3.3 Chromatin compaction and CRISPR/Cas9 editing efficiency

Individual *in vitro* studies from three different research groups have recently assessed whether Cas9 activity is inhibited by the presence of nucleosomes on target DNA. In 2015 Hinz *et al* demonstrated that CRISPR/Cas9 cleavage of its target sequence is inhibited when the latter is wrapped around a nucleosome. Positioning of the PAM site, rather than the protospacer, appeared to be the main determinant of the inhibitory effect (Hinz, Laughery and Wyrick, 2015). A few months later Isaac *et al* reached the same conclusion using similar methods but also showed that nucleosomal inhibition of CRISPR/Cas9 is alleviated by adding the RSC chromatin remodelling enzyme from *Saccharomyces cerevisiae* to the reaction to resemble physiological intracellular conditions (Isaac *et al.*, 2016). Similarly, the findings of Horlbeck *et al* using the Chd1 yeast chromatin remodeller also proved that factors promoting nucleosome dynamics increase cleavage efficiency, presumably by making otherwise occluded PAM sites transiently accessible support that the effect of nucleosome positioning in CRISPR/Cas9 activity is reversible (Horlbeck *et al.*, 2016). The overall conclusion from these three studies is that the presence of an immobile nucleosome over a PAM site hinders the activity of the CRISPR/Cas9 genome editing tool *in vitro*. However, chromatin *in vitro* is static compared to chromatin in the context of a living cell. As shown more recently, the editing potency of a certain sgRNA may be reduced in an *in vivo* setting compared to an *in vitro* assay (Yarrington *et al.*, 2018; Uusi-Mäkelä *et al*., 2018). Therefore, it is crucial to develop more physiologically relevant systems to investigate further what happens *in vivo*.

The first studies to report on the effect exerted by mammalian chromatin to CRISPR/Cas9 emerged in 2014. Genome-wide ChIP-seq binding profiles for catalytically dead Cas9 (dCas9) in mESCs and HEK293 cells showed that chromatin accessibility determines binding of the dCas9-sgRNA complex (Kuscu *et al.*, 2014; Wu *et al.*, 2014). In addition to the intended target site, dCas9 associated with several off-target sites with similar nucleotide

sequences. The likelihood of off-target binding was greater in DNase I hypersensitive sites (DHS), indicating that Cas9 also preferentially associates with open chromatin. This argument is further supported by the finding that CRISPR/Cas9 activity is highly correlated with H3K4me3, an epigenetic mark found at promoters of actively transcribed genes that is associated with open chromatin (Chari *et al.*, 2015). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) data from zebrafish embryos also verified that the higher the target accessibility, the higher the editing efficiency of CRISPR/Cas9 *in vivo* (Chen *et al.*, 2017; Uusi-Mäkelä *et al.*, 2018).

Single-molecule tracking techniques were used to determine the diffusion kinetics of dCas9 near heterochromatic nuclear domains. Halo-tagged dCas9 diffused at a significantly slower rate within nuclear domains occupied by the heterochromatin protein 1 (HP1) but was still able to penetrate heterochromatin and reach the sgRNA target site, as would also be expected for wt Cas9 (Knight *et al.*, 2015). Additional evidence from editing in yeast suggested that the nucleosome positioning occludes the association between the Cas9-sgRNA complex and its target site, regardless of the chromatin environment where the target site is embedded (Yarrington *et al.*, 2018).

Moreover, it is well established that sequence properties of the sgRNA, probably acting via secondary structures and effects on stability, have a major bearing on editing efficiency (Moreno-Mateos *et al.*, 2015; Jensen *et al.*, 2017). To assess the impact of chromatin modifications on editing efficiency, it is therefore desirable to assess the efficiency of editing on the same target site when it is embedded into different chromatin conformations.

Notable examples of such systems include artificially-engineered cell lines where differential chromatin compaction of a region coding for a fluorescent reporter is achieved through chemical induction. Chen *et al* made use of a HEK293-based system in which a KRAB-repressor protein is recruited to Tet Operator repeats flanking a Cas9 target site within a fluorescent reporter gene. Addition of doxycycline prevented recruitment of the repressor and correlated with a greater number of editing events as a consequence of

chromatin decompaction around the reporter construct (Chen *et al*., 2016). A follow up publication from the same group demonstrated that this principle also applied to Cas9n (Chen *et al.*, 2017). Similar effects have been observed by another group that utilised a different transgenic system in HEK293 cells, where compaction of the region containing the luciferase reporter gene could be induced upon addition of doxycycline. Surprisingly, after comparing editing efficiencies of multiple sgRNAs throughout the length of the luciferase reporter gene in open and closed chromatin, it became evident that chromatin did not influence all individual sgRNAs to an equal extent. Some sgRNAs showed no inhibition at all, whereas other sgRNAs situated at promoter-proximal positions among the luciferase reporter gene were profoundly inhibited (Daer *et al*., 2017). Taken together, these results suggest that chromatin compaction can be an important factor for the efficiency of the Cas9-sgRNA complex. However, it also raises the possibility that additional unknown factors may influence the way eukaryotic chromatin regulates mutagenesis by CRISPR/Cas9.

In 2015 Polstein *et al* provided the first evidence that dCas9 gained access to genomic regions that are protected from the transcription machinery. It was argued that while dCas9 associated with the target site determined by the sgRNA, the complex overcame the lack of accessibility by inducing chromatin remodelling of the region where the target site was embedded (Polstein *et al.*, 2015). Soon afterwards it was also demonstrated that dCas9 is sufficient to remodel chromatin to the extent of making it accessible for other Transcription Factors, whose binding motifs were previously occluded by nucleosomes (Barkal *et al*., 2016). The above findings are important points to consider about the behaviour of CRISPR/Cas9 *in vivo*. The fact that chromatin may interfere with CRISPR/Cas9 efficiency but Cas9 may remodel chromatin while searching for its target site creates a paradox. It also highlights the need for further investigation.

Despite the useful insight obtained from earlier studies comparing CRISPR/Cas9 editing efficiency using artificial differentially compacted reporter constructs (Chen *et al*., 2016; Daer *et al*., 2017) it is vital to examine whether the same effect can be observed on endogenous gene sequences

when present in their natural chromatin context. A recent review summarised what is currently known about how chromatin influences the efficiency of genome editing with CRISPR/Cas9 and emphasised the advantage of internally controlled systems that have been used to target the same DNA sequence when it is found in two different chromatin conformations within the same nuclear environment (Verkuijl and Rots, 2019).

To date, only two studies have been conducted using such systems. The first involved targeting of the *p161NK4a* locus in HCT116 cells and the second focused on targeting maternally-imprinted loci in mESCs (Fujita, Yuno and Fujii, 2016; Kallimasioti-Pazi *et al.*, 2018). In both cases the target sequences were present in two copies that were characterised by differential epigenetic chromatin modifications. In the first case the *p161NK4a* locus has two alleles that are decorated with DNA methylation and H3K9me2 in a mutually exclusive manner. Surprisingly, two out of the three employed sgRNAs were not affected by the allele-specific chromatin features of the target locus (Fujita, Yuno and Fujii, 2016). The findings of the second study are described in detail in the upcoming results' chapters of this thesis.

1.4 Mammalian genomic imprinting

1.4.1 Historical discovery, establishment and function of genomic imprinting

The notion that the two sets of chromosomes all mammals inherit from their parents have distinct functions during development was originally demonstrated by two different research groups in the early 1980s. Pronuclear transplantation was utilised to generate bimaternal and bipaternal mouse zygotes, which contain two copies of the maternal and paternal genomes in either case, respectively (McGrath and Solter, 1983; Surani and Barton, 1983). The resulting embryos were delivered to pseudo-pregnant female mice, resulting solely in short pregnancies that were characterised by early embryonic lethality shortly after implantation (McGrath and Solter, 1984a; Surani, Barton and Norris, 1984). Consecutive studies established that this drastic phenotype was owed to the presence of two genomic copies of the same parental origin, instead of other cytoplasm-related effects (Mann and Lovell-Badge, 1984; McGrath and Solter, 1984b).

A few years later, the discovery and characterisation of the first imprinted murine genes confirmed earlier speculations that a certain number of genes are expressed from only one of the two parental chromosomes throughout embryonic development (Cattanach and Kirk, 1985; Barlow *et al.*, 1991; Bartolomei, Zemel and Tilghman, 1991; DeChiara, Robertson and Efstratiadis, 1991; Ferguson-Smith *et al.*, 1991). Examples include the reciprocally imprinted insulin-like growth factor 2 (*Igf2*), which is maternally imprinted in mouse and human (DeChiara, Robertson and Efstratiadis, 1991; Giannoukakis *et al.*, 1993) and the insulin-like growth factor 2 receptor (*Igf2r*), which is paternally imprinted (Wutz *et al.*, 1997; Birger *et al.*, 1999).

To date, approximately 200 imprinted genes have been identified within the mouse and human genomes (Barlow, 2011; Ferguson-Smith, 2011;

Ferguson-Smith and Bourchis, 2018). The striking similarity of all imprinted genes is allele-specific DNA methylation, which is established de novo by essential DNA methyltransferases and acts as a platform for additional allele-specific epigenetic marks (Li, Beard and Jaenisch, 1993; Bourc'his *et al.*, 2001; Fournier *et al.*, 2002; Kaneda *et al.*, 2004). More specifically, the DNMT3L de novo DNA methyltransferase is recruited at regions where H3K4me3 is erased from the KDM1B histone demethylase (Ooi *et al.*, 2007; Ciccone *et al.*, 2009).

The allele-specific DNA methylation marks found on imprinted genes are termed 'imprints' and are established during gametogenesis in mammals. Meiotically-arrested oocytes acquire 'imprints' during oocytes growth (Obata *et al.*, 1998) whereas in males 'imprints' are laid prior to meiosis during testis development (Davis *et al.*, 2000; Ueda *et al.*, 2000). As illustrated in **Figure 1.6**, 'imprints' are stably maintained after fertilization throughout development, with the exception of the germline where they are reset and re-established in a gender-specific manner (Reik and Walter, 2001). There, the DNA methylation is completely erased and the 'imprints' are established de novo in a sex-specific manner during gonadal development (Barlow and Bartolomei, 2014).

Imprinting maintenance is achieved through the synergistic action of the pre-existing DNA methylation mark and additional factors that are recruited by it. The precise locations where the allele-specific DNA methylation is deposited near imprinted genes are called imprinting control regions (ICRs), which frequently contain repeats of the TGCCGC motif (Weaver *et al.*, 2010). When an ICR is methylated, the ZFP57 zinc finger protein associates with this motif and in turn recruits the KAP1 complex and DNA methyltransferases, which aid in maintenance of the local DNA methylation environment (Li *et al.*, 2008; Quenneville *et al.*, 2011). Similarly, PGC7/Stella also has an active role in DNA methylation maintenance of two imprinted genes in early embryogenesis through a yet undetermined mechanism (Nakamura *et al.*, 2007). Thus, 'imprints' are protected from the genome-wide waves of DNA demethylation that occur after fertilization and before embryonic implantation (Reik, Dean and Walter, 2001).

Genomic imprinting is essential for mammalian development (McGrath and Solter, 1984a; Surani, Barton and Norris, 1984; Cattanach, 1986). Embryonic stem cell lines with two copies of the paternal genome have been developed and injected to adult mice, where they cluster and form tumours (Mann *et al.*, 1990). In addition, it has been demonstrated that chimeric embryos that contain cells with two genomic copies from the same parent, develop clusters to retain these abnormal cells to the cerebral cortex (Allen *et al.*, 1995).

Most importantly, genomic imprinting is necessary to maintain a balance in expression between growth suppressor and growth enhancer genes in the placenta and other extraembryonic tissues. Paternally expressed genes are responsible for promoting embryonic growth, whereas maternally expressed genes suppress embryonic growth to guarantee the survival of the mother (Coan, Burton and Ferguson-Smith, 2005). For example, paternally expressed *Igf2* promotes growth in the fetus (Murphy *et al.*, 2015). Disruption of *Igf2* leads to growth deficiency in murine embryos (DeChiara, Efstratiadis and Robertsen, 1990), whereas knocking out *Igf2* further leads to disproportionate growth of the placenta (Coan *et al.*, 2008). On the contrary, maternally expressed *Igf2r* binds *Igf2*, leading to degradation of *Igf2* (Okas, Rozek and Czech, 1985; Zavorka *et al.*, 2016). Biallelic expression of *Igf2r* leads to a reduction in embryonic growth, which may persist into adult mice (Wutz *et al.*, 2001). Additional studies have linked the function of imprinted genes with postnatal processes, such as adaptation to feeding and social behaviour (Plagge *et al.*, 2004; Garfield *et al.*, 2011). Therefore, the expression of imprinted genes must be regulated appropriately for mammals to develop properly before and after birth.

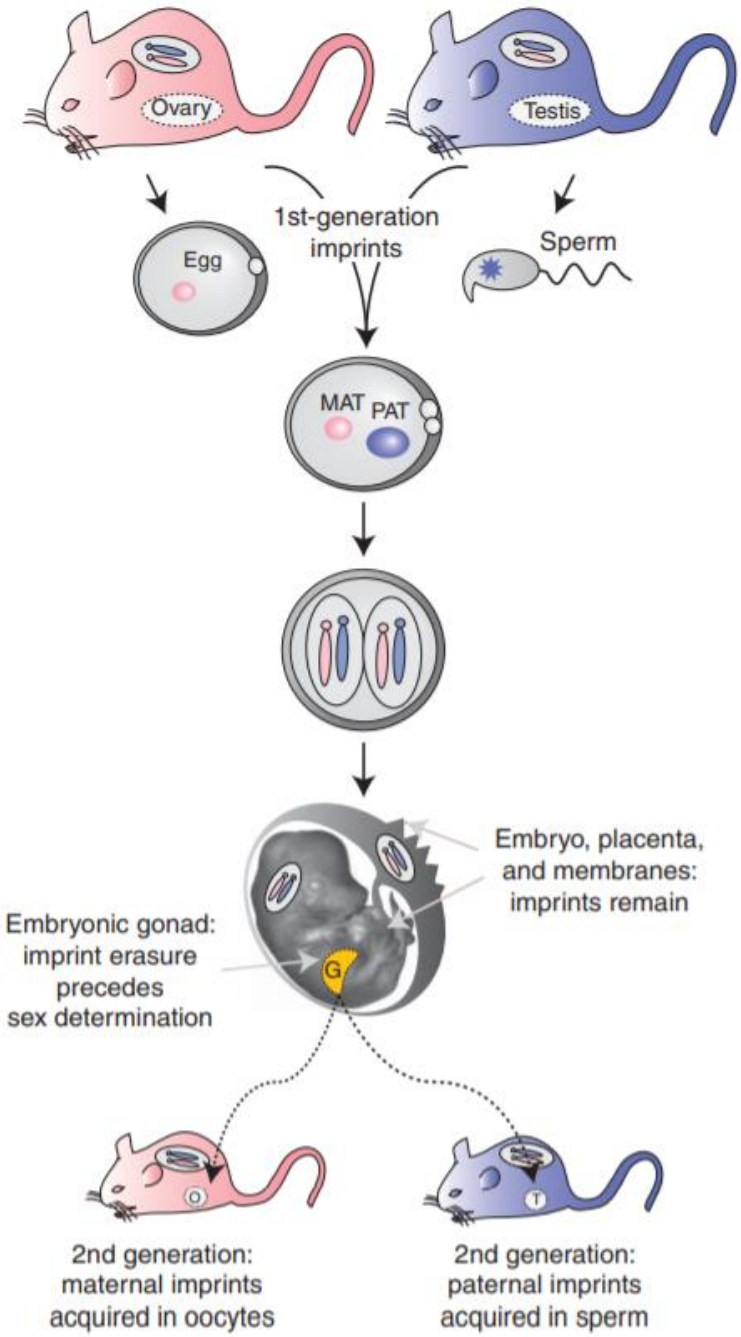


Figure 1.6 The life cycle of imprints in mice.

Imprint establishment occurs in female and male gametes, followed by maintenance throughout the embryonic and extraembryonic tissue. Prior to sex determination, imprints are erased from the embryonic gonad, where the germline will develop later. A second round of establishment occurs after the sex of the progeny is determined, when paternal and maternal imprints are acquired in sperm and eggs, respectively. **Adapted from Barlow and Bartolomei, 2014.**

1.4.2 Epigenetic properties and insightful contributions of imprinted genes

Imprinted genes must be expressed either from the maternal or paternal allele. With the exception of only a few, imprinted genes are found in clusters throughout the murine genome (Zwart *et al.*, 2001). Allele-specific silencing of imprinted genes is primarily determined by the ICR, which acquires differential DNA methylation during imprinting establishment (Edwards and Ferguson-Smith, 2007).

The expression of imprinted genes is regulated by 'imprints' and additional layers of epigenetic information. Chromatin immunoprecipitation (ChIP) experiments on mouse liver and brain tissue have first reported the presence of H3K9me3 on the methylated allele and H3K4me3 on the unmethylated allele of imprinted ICRs (Fournier *et al.*, 2002). Furthermore, the study of DNMT3L-depleted midgestation embryos that were derived from crossing female C57BL/6 and male JF1 mice provided useful insights into the link between DNA methylation and histone modifications at ICRs. ChIP for H3K9me3 and H4K20me3 revealed a dramatic loss of both of these repressive histone modifications at ICRs, which correlated with the absence of ICR methylation and biallelic expression of imprinted genes (Henckel *et al.*, 2009). This shows that repressive histone modifications require the germline-deposited CpG methylation mark on ICRs.

With respect to chromatin modifications, methylated sequences resemble repetitive and transposable elements that are naturally silenced in mammalian cells. Imprinted alleles and retrotransposons are transcriptionally silenced through a mechanism that involves the KAP1 complex recruiting HP1 (Rowe *et al.*, 2010). Additionally, both of the above are decorated with H3K9me3 and H4K20me3 to maintain a compacted status (Martens *et al.*, 2005; McEwen and Ferguson-Smith, 2010). Imprinted alleles therefore resemble constitutive heterochromatin (Regha *et al.*, 2007) (as shown in **Figure 1.7**).

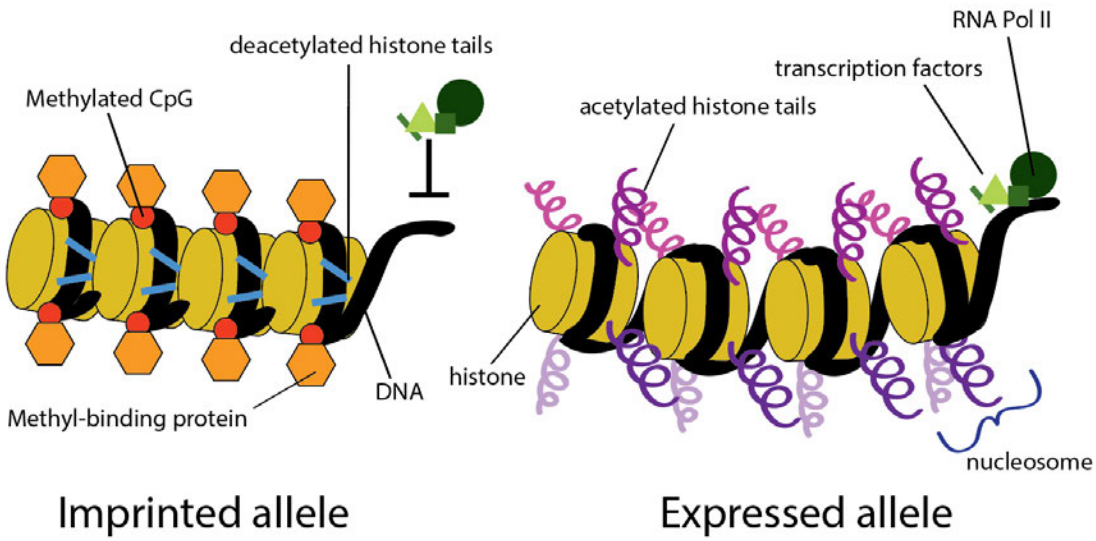


Figure 1.7 Allele-specific epigenetic properties of imprinted loci.

Representation of the epigenetic characteristic differences between alleles of imprinted genes. The ICR or promoter of a transcriptionally-silent imprinted allele is methylated and tightly wrapped around deacetylated histones (shown on the left). Proteins with Methyl-CpG-binding domains (MBDs) like MeCP2, MBD1 and MBD3 have been found to associate with hypoacetylated imprinted allelic regions (Fournier *et al.*, 2002). In contrast, the transcriptionally-active allele (shown on the right) does not undergo nucleosomal condensation due to the presence of acetylated histones. Additional histone modifications are mentioned in the text. **Modified from Reik and Walter, 2001.**

A final characteristic that is specific to the promoters of imprinted genes is differential accessibility and nucleosomal density. DNase I sensitivity assays have demontrated that the promoter of the unmethylated allele is sensitive to digestion, whereas the promoter of the methylated imprinted allele is inaccesible (Feil *et al.*, 1997).

To date, the majority of imprinted mammalian genes are controlled by maternally-methylated ICRs, where DNA methylation and repressive histone modifications work in synergy to silence the allele that is inherited from the mother (Ferguson-Smith and Bourchis, 2018). These genes are mostly found in large clusters but also, less frequenctly, as individual genes controlled by a single ICR (Ferguson-Smith, 2011). It would be preferable to assess the effect of chromatin modifications in ICRs from clusters of genes, but also in ICRs that regulate individual genes.

Maternally-imprinted ICRs have a high CG dinuclotide content and are highly influential for mammalian development. On the contrary, paternally-imprinted genes are usually located within intergenic regions, contain fewer CG dinucleotides and are more prone to imprinting erasure (Schulz *et al.*, 2010). The latter is further supported by a large-scale study in human pluripotent stem cells that demonstrated higher CG methylation loss at the ICRs of paternally-imprinted genes (Bar *et al.*, 2017).Therefore, for the purposes of my PhD, I have decided to only focus on maternally imprinted ICRs and target only maternally imprinted genes. Despite this decision, both types of imprinted genes have been used as model systems that led to important discoveries about epigenetic regulatory mechanisms of gene expression, like the relationship between DNA methylation and alternative polyadenylation on a maternally-imprinted gene (Wood *et al.*, 2008).

CTCF, or the CCCTC-binding factor, is a DNA binding protein that is highly conserved among vertebrates. This ubiquitous protein defines chromosomal domain boundaries and is mostly associated with transcriptional repression but also transcriptional activation (Filippova *et al.*, 1996; Vostrov and Quitschke, 1997; Bell, West and Felsenfeld, 1999). It was later established

that CTCF associates with the maternal unmethylated ICR of the *H19* paternally-imprinted gene, where it regulates the expression of the neighbouring *Igf2* maternally-imprinted gene by isolating *Igf2* from its cis regulatory enhancer (Bell and Felsenfeld, 2000; Hark *et al.*, 2000; Szabó *et al.*, 2000; Engel *et al.*, 2004). This is a case where genomic imprinting was repurposed into a system that led to discoveries about protein-mediated transcriptional regulation and chromosome topology (Murrell, Heeson and Reik, 2004).

Another major contribution to the understanding of transcriptional regulation was made through the elucidation of the role of long non-coding RNAs (lncRNAs) in imprinting maintenance. Imprinted gene clusters, such as *Igf2r* and *Cdkn1c/Kcnq1*, contain long non-coding transcripts that are specifically transcribed from the paternal chromosome and aid in transcriptional repression of nearby paternally-imprinted genes (Sleutels, Zwart and Barlow, 2002; Mancini-Dinardo *et al.*, 2006). Specifically, the *Air* lncRNA has a central role to the recruitment of the G9a H3K9 mono- and di-methyltransferase to the paternal chromosome (Nagano *et al.*, 2008). Similarly, *Kcnq1ot1* recruits the DNMT1 maintenance DNA methyltransferase (Mohammad *et al.*, 2010, 2012), G9a and PRC2 to block transcription in cis (Pandey *et al.*, 2008; Terranova *et al.*, 2008). A recent study even suggested a functional role for *Kcnq1ot1* that is not related with genomic imprinting, highlighting potential for further significant discoveries that can be made by repurposing genomic imprinting (Andresini *et al.*, 2019).

Genomic imprinting is arguably an excellent system to study epigenetic regulation of transcription. The presence of differential epigenetic modifications on two identical copies of the same underlying DNA sequence make imprinted genes a very attractive system for studying the effect of chromatin modifications on the efficiency of CRISPR/Cas9, as well as the repair of DSBs in the context of genome editing with an externally-provided single-stranded donor molecule.

1.5 Aims

The central goal of my PhD project was to understand how chromatin influences the efficiency of genome editing with the CRISPR/Cas9 programmable nuclease and identify which DNA DSB repair pathway is favoured by different chromatin contexts. The specific aims of my project were:

1) To establish a system using genomic imprinting to study the effect of chromatin modifications on genome editing with CRISPR/Cas9.
2) To investigate how additional factors may influence the effect of chromatin on genome editing with CRISPR/Cas9.
3) To assess how chromatin modifications affect off-target site mutagenesis by CRISPR/Cas9.
4) To determine the effect of chromatin on mutagenic repair outcomes during genome editing with CRISPR/Cas9.

Chapter 2

Materials and Methods

2.1 Molecular cloning

2.1.1 Design and preparation of sgRNAs

Target regions near promoters and within imprinted gene bodies were selected based on publicly available DHS data in the UCSC Genome Browser. SNP information for sgRNA selection was obtained from the National Institute of Genetics (NIG) mouse genome database (http://molossinus.nig.ac.jp/msmdb/index.jsp). The National Institute of Genetics, in Mishima, Shizuoka, Japan, is where the Japanese funky 1 mouse (JF1) was first sequenced (Takada *et al.*, 2013). The CRISPR sgRNA designer for the Broad Institute (Doench *et al.*, 2014) was used to select sgRNAs (http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). Guides with GGG PAM sequences, high On-Target Efficacy Scores and close proximity to a SNP in the murine JF1 genome were selected for cloning.

Complementary oligonucleotides, consisting of the 20 nucleotide-long protospacer sequence and a 4 nucleotide-long overhang, were ordered from Integrated DNA Technologies. After pairwise phosphorylation with T4 Polynucloetide kinase (NEB, M0201S), the phosphorylated complementary oligonucleotide pairs were heated at 95 °C for 5 minutes and allowed to gradually cool down and anneal. All sgRNA oligonucleotide sequences are listed in **Appendix 1.1**.

2.1.2 Cloning sgRNAs in mammalian CRISPR vectors

The universally available protocol by Feng Zhang's lab was used to clone all guides in the GFP (px458, Addgene Plasmid #48138) and Puromycin resistance (px459v2*,* Addgene Plasmid #62988) vectors (Ran, Hsu, Wright, *et al.*, 2013). In addition, gInpp5f_v2, gImpact1, gKvDMR#1 and gKvDMR#3

were also cloned into a vector coding a 3xFLAG-dCas9 (pX330A_dCas9-1x2, Addgene Plasmid #63596) for dCas9 ChIP. All vectors were digested with FastDigest BpiI (ThermoFisher Scientific, FD1014) and dephosphorylated with FastAP (ThermoFisher Scientific, EF0654). Annealed complexes of phosphorylated sgRNA oligonucleotides were ligated to all vectors with the T4 DNA Ligase (NEB, M0202S).

2.1.3 Design of ssODNs

All ssODNs were designed to span the sgRNA target site and contain a mutation that abolishes the PAM sequence and generates a restriction enzyme site. Detection of HDR events is possible through identification of the pre-engineered mutation in the NGS reads or by enzymatic digestion. The ssODNs used in this study were 150bp long and did not overlap with any SNPs, to avoid introducing bias to the overall mutagenesis result. All ssODN sequences were ordered as custom single-stranded ultramers from Integrated DNA Technologies, a full record is provided in **Appendix 1.2**.

2.2 Cell Culture

2.2.1 Hybrid mESC culture and maintenance

BxJ and JxB hybrid mESCs were derived as extensively described in section 3.2, cultured in ESGRO Complete PLUS Clonal Grade Medium (Millipore, SF001-500P) and passaged once every two days. Cells were seeded in standard plastic cell culture flasks and plates that were coated with

0.2% Gelatin (Sigma, G1393-100ML) solution for 2 minutes immediately before seeding.

2.2.2 Transfection and puromycin selection

Cells, with passages ranging from 6 to 12 unless otherwise specified, were plated in 6-well plates. The amount of 0.3 million cells was seeded in each well of a 6-well plate and transfected with 3ug plasmid DNA and 2 µl of 10uM ssODN using Lipofectamine 2000 (ThermoFisher Scientific, 11668019). Transfections were performed 24 hours after seeding the indicated amount of cells. Unless otherwise mentioned, two sgRNAs and ssODNs were transfected simultaneously. The effect of co-transfection on editing efficiency is not known. Additional details from the dCas9 cross-linked ChIP transfections are mentioned in section **2.4.2**. For all transfections, lipofectamine-containing media was removed five hours post-transfection to avoid cytotoxic effects. Twenty-four hours post-transfection cells were subjected to 1.5ug/ml Puromycin-containing media (ThermoFisher Scientific, A1113803) for twenty-four hours. Longer Puromycin selection timings resulted in extensive cell death. Recovery after puromycin selection required maintaining the cells in culture for 48 hours, leading to cell harvesting after puromycin selection at 4 days (or 96 hours) after transfection.

Due to the limited time of reagent exposure during early time points, time course experiments were performed without puromycin selection. In addition, the experiments shown in section **4.3** of Chapter 4 were performed with slight deviation from the general protocol. At 24 hours after transfection, cells were FACS sorted according to GFP fluorescence levels and maintained as distinct populations. Half of the cells from each population were immediately harvested and genomic DNA was extracted for downstream processing. The remaining half was re-seeded and maintained in culture for a further 72 hours. Transfection efficiency ranged between 30% and 40%, as determined by GFP

fluorescence during the experiments described in section 4.3. Alternatively, during experiments with puromycin selection, cell viability was used as a quality control measure. Cell density was observed 24 hours after puromycin addition, if confluency was below 40% then the experiment was discarded.

2.2.3 Hybrid mESC 2i culture with vitamin C supplement

A vial with BxJ passage 6 cells was thawed and plated in ESGRO. Cells were maintained in 1i for 3 passages (always 1:10 split). A day after the third passage was seeded, the cells settled down. ESGRO was removed, the cells were washed with PBS and supplied with 2i media (PD0325901 and CHIR99021 inhibitors for MEK and GSK3, respectively). 2i media was supplied daily and the cells were passaged (1:10) every two days until they reached passage 20. After the cells settled down 1 day after plating passage 20, the 2i media was replaced with 2i with L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate (vitamin C, Sigma, A8960) at a concentration of 100ug/ml. 2i media with 100ug/ml vitamin C was also changed daily and cells spent 4 days under these conditions. They were passaged twice during this time, reaching passage 22. The culture medium was shifted back to ESGRO 1 day after plating passage 22 and then the cells were trypsinised and plated for transfections in 6 well plates as usual, in ESGRO. Tranfections in BxJ passage 23 were performed as previously described with Lipofectamine 2000. Cells were harvested at 16h and 96h after transfection in the absence of Puromycin selection. A complete recipe for the 2i culture medium is provided in **Appendix 2**.

2.2.4 Genomic DNA extraction following transfections

Cells were collected at indicated times after transfection for separate experiments. Harvested cells were spun down and resuspended in 200 µl PBS. Genomic DNA was isolated with the DNeasy Blood & Tissue kit by Qiagen. Samples were quantified on a NanoDrop Spectrophotometer from ThermoFisher Scientific.

2.2.5 Fluorescence activated cell sorting (FACS)

Flow cytometric analysis of cells transfected with px458 was conducted on a Becton Dickinson FACSAria II instrument by the FACS technicians of the MRC HGU, Elisabeth Freyer and Stacey Thomson. Data analysis was performed in BD FACSDiva (Becton Dickson, Version 6.1.2) also by Elisabeth Freyer and Stacey Thomson.

2.3 Preparation and validation of NGS libraries

2.3.1 Target locus amplification and labelling

Amplicons for all target regions were designed to cover the sgRNA target site and at least one allele-specific SNP. A two-round PCR amplification strategy was employed, similar to the ‘16S Metagenomic Sequencing Library Preparation Protocol’ by Illumina. Phusion High-Fidelity DNA Polymerase (NEB, M0530S) was used instead of the recommended 2x KAPA HiFi HotStart ReadyMix. The PCR recipe was adjusted according to the Phusion High-Fidelity DNA Polymerase online protocol by NEB (https://www.neb.com/protocols/1/01/01/pcr-protocol-m0530).

Primers were designed to contain adaptor and barcode sequences necessary for multiplexed high-throughput amplicon NGS. Initially, 50ng of genomic DNA sample was amplified for 25 cycles with region-specific primers for every locus. Primers for the first amplification step contained 5´ extensions with a random hexamer, binding sites for Illumina sequencing primers, and binding sites for the primers of the second round of amplification. First round amplicons were purified using AMPure XP beads (Beckman Coulter, A63882) as described in the '16S Metagenomic Sequencing Library Preparation Protocol' by Illumina and eluted in 50ul. One fifth of the eluate (10ul) was used for a second round of amplification for 8 cycles using primers with Illumina indices (i5 and i7). Second round amplicons were also purified using AMPure XP beads and eluted in 25ul. All first round (region-specific) and second round (universal) primer sequences are provided in **Appendix 3.1** and **Appendix 3.3**, respectively.

Concentration, size and sample purity of the generated NGS libraries was examined by running 1 µl of each library on an Agilent Bioanalyser High Sensitivity chip. Stock dilutions of amplicon libraries passing quality control were prepared individually and pooled at equimolar ratio at a final concentration of 4nM. NGS was performed on the MiSeq Illumina platform to generate 150bp paired-end reads.

2.3.2 Bisulfite library preparation

Genomic DNA samples were converted with the EZ DNA Methylation-Gold Kit by Zymoresearch. The amount recommended in the protocol (500ng) was used per conversion reaction. Region-specific primers for bisulfite converted DNA were designed on Methprimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi). As for non-converted samples, 5´ extensions with a random hexamer, binding sites for Illumina sequencing primers, and binding sites for the second-round primers were added to the region-specific

primers for the first round of amplification. Bisulfite modified genomic DNA (100ng) was amplified across regions spanning the sgRNA targeting sites to cover 6-10 CpG dinucleotides, as it was not possible to include a SNP for every target region. All first round bisulfite primer sequences and the number of CpGs for individual amplicons are provided in **Appendix 3.2**. The first round of amplification required 35 cycles and different annealing temperatures for different region-specific primer pairs. To obtain a decent amplicon amount, the 2XGoTaq Green Master Mix (Promega, M7123) was used for both amplification rounds. All region-specific primer sequences and annealing temperatures are listed in **Appendix 3.2**.

2.4 Analytical and functional assays

2.4.1 Surveyor assay

Edited genomic DNA samples were subjected to the Surveyor assay to assess the presence of NHEJ-related events and derive information about overall genome editing efficiency levels prior to NGS. Initially, PCR amplicons covering a wide region of 250 nucleotides on either side of the cleavage site, were generated using 200ng of each sample. Following 30 cycles of amplification, amplicons were purified using the QIAquick PCR Purification Kit by Qiagen. Purified amplicons were quantified on a NanoDrop Spectrophotometer from ThermoFisher Scientific.

Subsequent steps, such as DNA heteroduplex formation, Surveyor nuclease S digestion and visualisation, were performed as in a widely-used published protocol (Ran *et* al., 2013). Digestions reactions were performed with the Surveyor Mutation Detection Kit (Integrated DNA Technologies, 706020) and cleavage products were visualised in a 2% agarose gel stained with the sensitive SYBR Gold Nucleic Acid Gel Stain (ThermoFisher Scientific, S11494).

2.4.2 DNase I accessibility assay

A published protocol (McArthur, Gerum and Stamatoyannopoulos, 2001) was modified to fulfil the purposes of our experiments. The amount of 20 million mESCs from each of the two hybrid lines was harvested and resuspended in 5ml of Buffer A. Nuclei were isolated following cell lysis with 0.5 % (v/v) NP40 and centrifugation at 2000g for 5 minutes and resuspended in 1ml of digestion buffer. The amount of 100 µl of nuclei were separately digested with 0–60 units of DNase I at 37°C for 5 minutes and discontinued after adding 1ml of stop buffer. Protein degradation was performed with 2ug Proteinase K treatment at 55°C for 16 hours. Phenol/chloroform extraction and ethanol precipitation were performed to recover the DNA, which was resuspended in TE buffer. The Qubit high sensitivity assay (ThermoFisher Scientific, Q32854) was used to measure the concentration of the resulting DNA samples. Regions of interest including one allele-informative SNP were amplified from the resulting DNA samples for 30 cycles. AMPure XP beads were used to purify amplicons, which were sequenced across the SNP with traditional Sanger sequencing. All relevant primer sequences are provided in **Appendix 4.2**. All buffer recipes for this assay are provided in **Appendix 5.1**. This assay was exclusively performed by Gillian Taylor.

2.4.3 Native histone ChIP

The amount of 10 million mESCs from each of the two hybrid lines was harvested and resuspended in 10ml chilled PBS. Cells were pelleted at 500g for 5 minutes at 4˚C and resuspended in 1ml chilled NBA buffer. Cell lysis was performed on ice for 5 minutes with the addition of 1ml of NBB buffer. Nuclei were collected after centrifugation of cell lysates at 1000g for 5 minutes at 4˚C and resuspended in 200 µl of NBR buffer. Nuclei were pelleted after a second centrifugation step of 5 minutes at 4˚C and resuspended in 600 µl NBR buffer, where 10 µl of 10mg/ml PureLink RNaseA (Invitrogen, 12091039) was added for 5 minutes for total RNA degradation at room temperature. A total of 40 Boehringer units of MNAse were used for an incubation at 20˚C for 10 minutes. Samples were mixed by pipetting when MNAse was added, as well as 5 minutes later, during the incubation. Addition of 600 µl MNase stop buffer stopped the digestion process and chromatin samples were spun down at 13,000rpm for 5 minutes at 4˚C.

In parallel, Protein A Dynabeads (Invitrogen, 10002D) were pre-washed in Blocking buffer and mixed and incubated with 2.5ug of antibody (per sample) on a rotating wheel for 2 hours at 4˚C. The amount of 40 µl of beads in suspension was used for every sample. Antibodies against H3K9me3 (07-442, batch 2664282) and H4K20me3 (07-643, batch 2586586) were purchased from Millipore. Following conjugation, the beads were washed with 200 µl Blocking buffer. At this point, 100 µl of the previously prepared chromatin samples was kept separately to be used as 10% Input for later ChIP quantification by qPCR. For the immunoprecipitation, 1ml of the previously prepared chromatin samples was mixed with the immune-conjugated beads and 5 µl of 5mg/ml BSA solution. Chromatin samples and beads were incubated on a rotating wheel for 3 hours at 4˚C.

Immunoprecipitated chromatin and beads were subsequently washed thrice with 1ml ChIP-W1 buffer for 10 minutes at 4˚C on a rotating wheel and once with TE buffer without nutation. Chromatin was released from the beads

following addition of 100 µl Elution buffer, vortexing and an incubation at 37˚C for 1 hour at 700rpm. The amount of 7µl of 2M Tris-HCl pH6.8 solution was added to adjust the pH value of the samples to 8. A short centrifugation step was necessary to discard the beads. The ChIP supernatant was kept and subsequently processed along with the 10% Input sample. Histones were degraded after a 1 hour treatment with 20ug of Proteinase K treatment at 55°C. All ChIP and 10% Input DNA samples were purified with the QIAquick PCR Purification Kit by Qiagen. Chromatin modification mapping at the chosen imprinted regions in the two hybrid mESC lines was exclusively performed by Gillian Taylor. Buffer recipes are provided in **Appendix 5.2**.

2.4.4 Cross-linked ChIP

Transfections for this experiment were slightly different from the procedure described in section **2.2.2**. Hybrid mESCs were seeded in 10cm dishes and transfected with 24ug of total plasmid DNA with Lipofectamine 2000. Four plasmids coding a 3XFLAG-dCas9 and four sgRNAs (gImpact1, gInpp5f_v2, gKvDMR#1 and gKvDMR#3) were co-transfected without any of their corresponding ssODNs. Transfected cells were harvested at the relevant time points without any selection, counted and resuspended in 10ml PBS. The amount of 10 million cells was subjected to cross-linking with 5ml chilled PBS containing 1% formaldehyde (ThermoFisher Scientific, 28906) for 10 minutes at room temperature with nutation. The cross-linking process was stopped with the addition of 1M glycine solution to a final concentration of 0.125M. Following cross-linking, cells were incubated for 5 minutes at room temperature with nutation, pelleted at 2000rpm for 5 minutes at 4˚C and washed with 10ml chilled PBS. Cell lysis was performed with 5ml of Farnham lysis buffer and nuclei were pelleted at 2000rpm for 5 minutes 4˚C prior to storage at -80˚C. After all time points were collected, they were processed in parallel. Once thawed, pelleted nuclei were resuspended once more in 5ml Farnham lysis buffer. After pelleting again at 2000rpm for 5 minutes 4˚C, nuclei were

resuspended in 1ml RIPA buffer. Sonication was then performed by 3 cycles of 30s on followed by 30s off on the Soniprep 150 (power = 4 amplitude microns) to prevent preferential shearing of the euchromatic allele. Fragment size ranged from 1-10kb, with the majority of fragments being around 3-5kb. Chromatin samples were spun down at 13,000rpm for 5 minutes at 4˚C and kept on ice.

In parallel, Protein G Dynabeads (Invitrogen, 10004D) were pre-washed in Blocking buffer and mixed and incubated with 5ug of antibody (per sample) on a rotating wheel for 2 hours at 4˚C. The amount of 40 µl of beads in suspension was used for every sample. A monoclonal antibody against FLAG (Sigma, F1804) was used to pull down 3X-FLAG-dCas9. Following conjugation, the beads were washed with 200 µl Blocking buffer. At this point, 100 µl of the previously prepared chromatin samples was kept separately to be used as 10% Input for later ChIP quantification by qPCR. For the immunoprecipitation, 1ml of the previously prepared chromatin samples was mixed with the immune-conjugated beads. Chromatin samples and beads were incubated on a rotating wheel for 18 hours at 4˚C.

Immunoprecipitated chromatin and beads were subsequently washed five times with 1ml LiCl Wash Buffer buffer for 5 minutes at 4˚C on a rotating wheel and once with TE buffer without nutation. Chromatin was released from the beads following addition of 200 µl Elution buffer, vortexing and an incubation at 65˚C for 1 hour at 1,000rpm. A short centrifugation step was necessary to discard the beads. The ChIP supernatant was kept and subsequently processed along with the 10% Input sample. All samples were incubated with 20ug PureLink RNaseA for 30 minutes at 37˚C to degrade total RNA. Proteins were degraded after a 5 hour treatment with 40ug of proteinase K treatment at 65°C. All ChIP and 10% Input DNA samples were purified with the QIAquick PCR Purification Kit by Qiagen. Buffer recipes are provided in **Appendix 5.3**.

2.4.5 qPCR and sequencing

Histone occupancy and dCas9 binding patterns were quantified with qPCR. Firstly, a standard curve for ChIP samples was generated from serial fivefold dilutions of the 10% Input samples of total MNase digested native chromatin and total sonicated chromatin. SYBR Select mastermix (Applied Biosystems, 4472908) was utilised to conduct reactions in triplicate on a LightCycler 480 II (Roche). Total reaction volumes were 20ul, including 0.5 µl of neat 10% Input or ChIP DNA.

For native histone ChIP, sequencing amplicons containing a SNP were prepared with the 2XGoTaq Green Master Mix. AMPure XP beads were used to purify amplicons that were subsequently sequenced with traditional Sanger sequencing. Alternatively, cross-linked ChIP material was used to generate NGS libraries as described in section **2.3.1**. All qPCR and sequencing primers are listed in **Appendix 4.1**.

2.5 Computational methods summary

2.5.1 Allele-specific data

MiSeq reads from all samples, including bisulfite-converted, were de-multiplexed. Duplicate read pairs were removed by FastUniq v1.1 (Xu *et al.*, 2012), and TrimGalore v0.4.1 was used to trim Illumina adaptor sequences (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

BWA v0.7.12 (Li and Durbin, 2010) was used to align processed read pairs to the mouse genome (build GRCm38). Information about the targeted genomic regions from each experiment allowed mapping read pairs to the

relevant genomic regions. Polymorphic SNP information from the National Intsitute of Genetics (NIG) mouse genome database (http://molossinus.nig.ac.jp/msmdb/index.jsp) was used to assign read pairs to the C57BL/6J or JF1 chromosome, to determine allele specificity. To extract DSB repair information read pairs were interrogated near the cleavage site of every corresponding sgRNA. The presence of InDels within 10 nucleotides of the sgRNA cleavage site indicated the presence of an NHEJ event. Notably, the presence of pre-engineered mutations from designed ssODNs indicated the presence of an HDR event. The absence of InDels and the pre-engineered mutation was interpreted to indicate the absence of genome editing on the corresponding read pair, as we were unable to rule out 'scarless' c-NHEJ upon cleavage in our system. The length of insertions and deletions from all NHEJ read pairs were quantified with a custom Perl script by Alison Meynert. The following link in the University of Edinburgh's Gitlab contains all of Alison's scripts along with some instructions and examples of parameter files (https://git.ecdf.ed.ac.uk/igmmbioinformatics/crispr-allele-specific-effects).
Analysis of all data was systematically performed by Alison Meynert and Tracy Ballinger.

2.5.2 Bisulfite data

Bismark v0.16.3 (Krueger and Andrews, 2011) and Bowtie v2.2.6 (Langmead and Salzberg, 2012) were used to align processed read pairs to the bisulfite conversion of the indexed mouse genome (build GRCm38). Read pairs that could not be aligned had to be broken down to single-end reads, which could then be aligned from each side of the target region separately. After merging the two alignments for single-end reads and the alignment for intact read pairs, information about the targeted genomic regions allowed mapping read pairs to the relevant genomic regions. A custom Perl script by Alison Meynert was used to count the number of methylated CpGs in all reads. The following link in the University of Edinburgh's Gitlab contains all of Alison's

scripts along with some instructions and examples of parameter files (https://git.ecdf.ed.ac.uk/igmmbioinformatics/crispr-allele-specific-effects).

Since most bisulfite amplicons did not cover allele-informative SNPs allele-specificity information was not determined. Instead, only the total methylation level of both alleles is reported. Analysis of all data was systematically performed by Alison Meynert and Tracy Ballinger.

Chapter 3

Repurposing genomic imprinting to study the effect of chromatin modifications on CRISPR/Cas9 mutagenesis

3.1 Introduction

Genome editing with CRISPR/Cas9 in mammalian cells is remarkably efficient, and is very likely to play an important role in the clinic in the foreseeable future. However, the natural function of the CRISPR locus is to degrade invading nucleic acids and render prokaryotes immune to bacteriophages (Barrangou *et al.*, 2007). The mechanism by which this small molecule manages to find its target site within highly compacted eukaryotic genomes is still not fully understood and merits further investigation. For example, we must improve our understanding of how CRISPR/Cas9 encounters eukaryotic chromatin before we start using it as a therapeutic tool towards treating human disease.

In the past, a series of studies have reported that nucleosome positioning has a negative impact on CRISPR/Cas9 activity *in vitro*, especially when the PAM site is fully protected by a nucleosome (Hinz, Laughery and Wyrick, 2015; Horlbeck *et al*., 2016; Isaac *et al*., 2016). It is logical to expect though that genome editing experiments *in vivo* are more likely to provide us with evidence of what would happen in a more clinically-relevant scenario.

In addition to the aforementioned studies, chromatin compaction has been shown to reduce the kinetics of Cas9 diffusion within live cells (Knight *et al.*, 2015). Several research groups have used a set of different cell-based and animal-based systems to demonstrate that chromatin compaction and epigenetic modifications reduce CRISPR/Cas9 mutagenesis (Chen *et al*., 2016; Fujita, Yuno and Fujii, 2016; Daer *et al*., 2017; Jensen *et al*., 2017; X. Chen *et al*., 2017; Uusi-Mäkelä *et al*., 2018; Yarrington *et al*., 2018). Nevertheless, it is speculated that Cas9 can remodel chromatin (Polstein *et al*., 2015; Barkal *et al*., 2016). This creates a paradox, and requires systematic quantification of genome editing in mammalian cells in a natural system, with minimal perturbations compared to previously used systems, where accessible and inaccessible chromatin states can be compared with higher fidelity.

In this chapter, I will describe how we repurposed genomic imprinting as a natural system to study the effect of chromatin modifications on CRISPR/Cas9 mutagenesis. This system was selected due to the advantage it provides by allowing us to assess genome editing efficiency when target sites of the same underlying DNA sequence are present within two different conformations within the same nuclear environment.

The work reported in this chapter was performed in collaboration with a number of different people. The hybrid mESCs used in this study were derived and provided to the Wood lab by Robert Feil's group. Gillian Taylor and Keerthi Chathoth and I characterised the epigenetic landscape of the intended target sites within endogenous imprinted loci, in an allele-specific manner. Alison Meynert set up the computational pipeline for using the Next generation sequencing data that I generated to quantify mutagenesis frequencies in an allele-specific manner.

3.2 Deriving the reciprocal hybrid mouse embryonic stem cell lines

Our goal was to compare CRISPR mutagenesis on the same DNA sequence in two different chromatin conformations within the same nucleus. We therefore decided to target maternally-imprinted loci. As described in **Chapter 1.4**, we would expect the maternal allele of these loci to be transcriptionally silent and compacted into heterochromatin. In addition, we would also expect the paternal allele to be transcriptionally active and euchromatic (Fitzpatrick, Soloway and Higgins, 2002).

Mouse embryonic stem cells have been used extensively for performing gene targeting (Capecchi, 2005). Some of the most important features of mESCs are their pluripotency, but also that they can be easily transfected and have a high uptake of exogenous repair templates due to efficient HDR (Glaser, Anastassiadis and Stewart, 2005).

In the past, Ildem Sanli and Sébastien Lalevée from Robert Feil's group had derived two male hybrid mouse cell lines by crossing *Mus musculus* C57BL/6J (B) and *Mus musculus molossimus* JF1 (J) mice in a reciprocal manner (Sanli *et al*., 2018) as shown and described in **Figure 3.1**. Notably, the BxJ line was generated by crossing a female B6 mouse with a male JF1 mouse. In addition, the JxB line was generated by crossing a JF1 female mouse with a male B6 mouse. Both lines were derived under serum-free conditions, specifically 1i culture media with only GSK3 inhibitor (**Figure 3.1**), to maintain pluripotency and methylation imprints (Wray *et al.*, 2011; Leitch *et al.*, 2013; Kota *et al.*, 2014) and subsequently karyotyped (Nagy *et al.*, 2008). The 1i culture media was chosen over 2i culture media (with MEK and GSK3 inhibitors) since the latter has been shown to affect the maintenance of CpG methylation by impairing the function of DNMT1 (Choi *et al.*, 2017; Yagi *et al.*, 2017). Male lines were selected over female ones, due to the genome-wide reduction of DNA methylation observed in female mouse ES cells (Zvetkova *et al.*, 2005). This is owed to the presence of two active X chromosomes and can be restored by eviction of one of the two active X chromosomes, leading to chromosomal instability in female mouse ES cell lines (George *et al.*, 2007). The reason why we opted for using hybrid lines was the presence of single-nucleotide polymorphisms (SNPs), meaning that some nucleotides will be distinct between the B and J genomes (Takada *et al.*, 2013). This would allow us to distinguish between the two imprinted alleles in our system and quantify mutagenesis by CRISPR/Cas9 in the two different localised chromatin environments.

Targeting experiments were performed in parallel for both reciprocal hybrid lines. This would be necessary to demonstrate that any allele-specific editing preference we observe is owed to the allele-specific epigenetic

chromatin modifications instead of the genetic background of the parental mouse strains. Before I began using CRISPR/Cas9 to target maternally imprinted loci, Gillian Taylor and Keerthi Chathoth selected and validated 3 endogenous maternally-imprinted loci. The data presented in the following section was generated by Gillian and Keerthi, who characterised the allele-specific chromatin states of the intended target loci and confirmed the presence of the relevant imprinting marks.

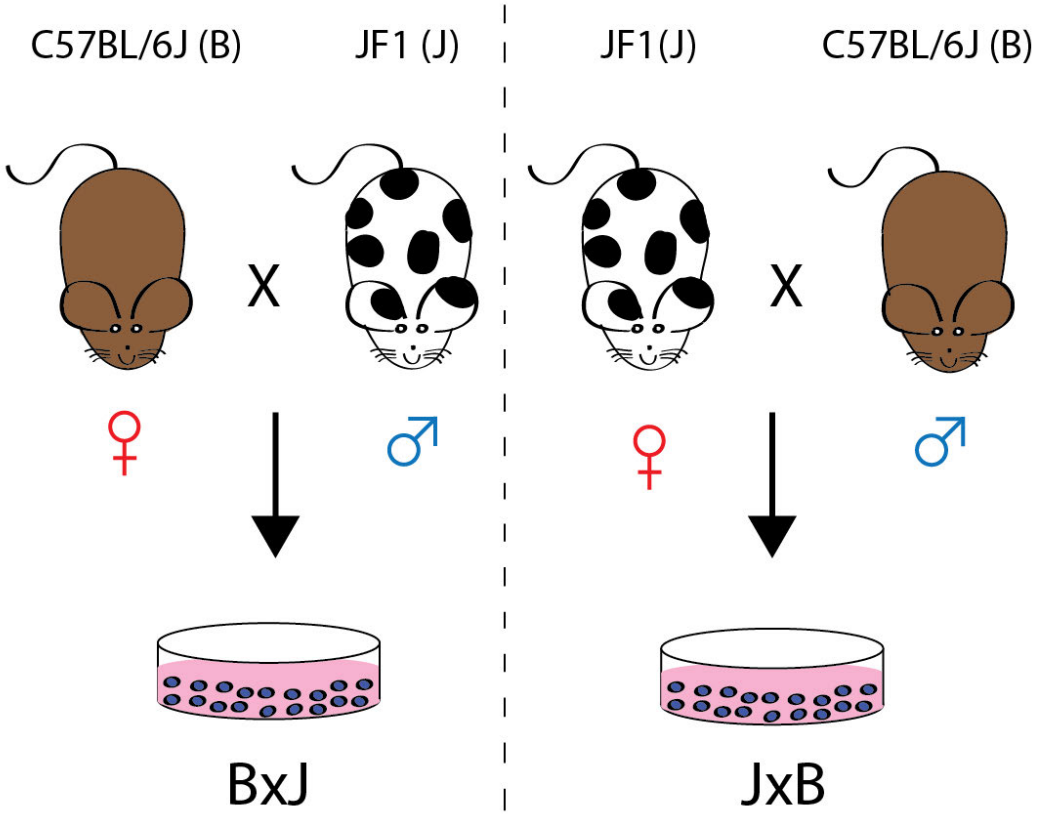


Figure 3.1 Hybrid mESC derivation process.

C57BL/6J (B) and JF1 (J) mice were crossed as described in the text. The initial of maternal genome of origin preceded the paternal initial in the given name of both reciprocal lines. ES cells were isolated from the inner-cell mass of blastocysts in 2i culture medium supplemented with LIF and in the absence of serum (Sanli *et al.*, 2018). The 2i culture medium contained inhibitors for MEK (PD0325901) and GSK3 (CHIR99021). After derivation, the cells were maintained in 1i culture medium. Unless otherwise stated, both hybrid mESC lines were cultured in the same 1i culture medium containing a GSK3 inhibitor (CHIR99021) for all downstream applications.

3.3 Epigenetic characterisation of target loci

The first selected locus was *KvDMR1*, which is a CpG island that is methylated only on the maternal chromosome 7 in mice (Smilinich *et al.*, 1999). *KvDMR1* is part of a large well-studied cluster of maternally imprinted genes and will be the locus that is used as an example of a cluster, as previously mentioned in **Chapter 1.4.2**. To avoid confusion with the allocated sgRNA names, the targeted region will be referred to as *KvDMR*. Keerthi Chathoth initially designed sgKvDMR#1 to target the ICR of the locus. Consequently, I designed sgKvDMR#2 and sgKvDMR#3 to target the promoter of the *KvDMR* region as described in **Chapter 2.1.1**. The exact position of each of the three sgRNA binding sites is shown in **Figure 3.2.A**, along with data for H3K9me3 ChIP and DNase I–seq in mESCs from the ENCODE project that were obtained through the UCSC genome browser.

We were initially interested to confirm whether the hybrid mESCs had allele-specific differences in chromatin accessibility at the *KvDMR* region, as previously reported (Feil *et al.*, 1997). A series of digestions with increasing units DNase I on whole nuclei at 37 °C for 5 minutes were performed by Gillian Taylor. The digested material was used to prepare PCR amplicons that were subjected to Sanger sequencing. The positions of the corresponding Sanger amplicons are shown in **Figure 3.2.A**. In **Figure 3.2.B** we focused on SNPs for either target site and observed two peaks on the chromatogram for zero units of DNase I. However, as the concentration of DNase I increases, we observe that the SNP from the paternal allele disappears while the SNP from maternal allele persists. The effect is more profound at the promoter region (sgKvDR#2&3), where the paternal SNP is lost with as little as 100 units of DNase I. We thus concluded that the paternal allele of the *KvDMR* region is more accessible than the maternal allele.

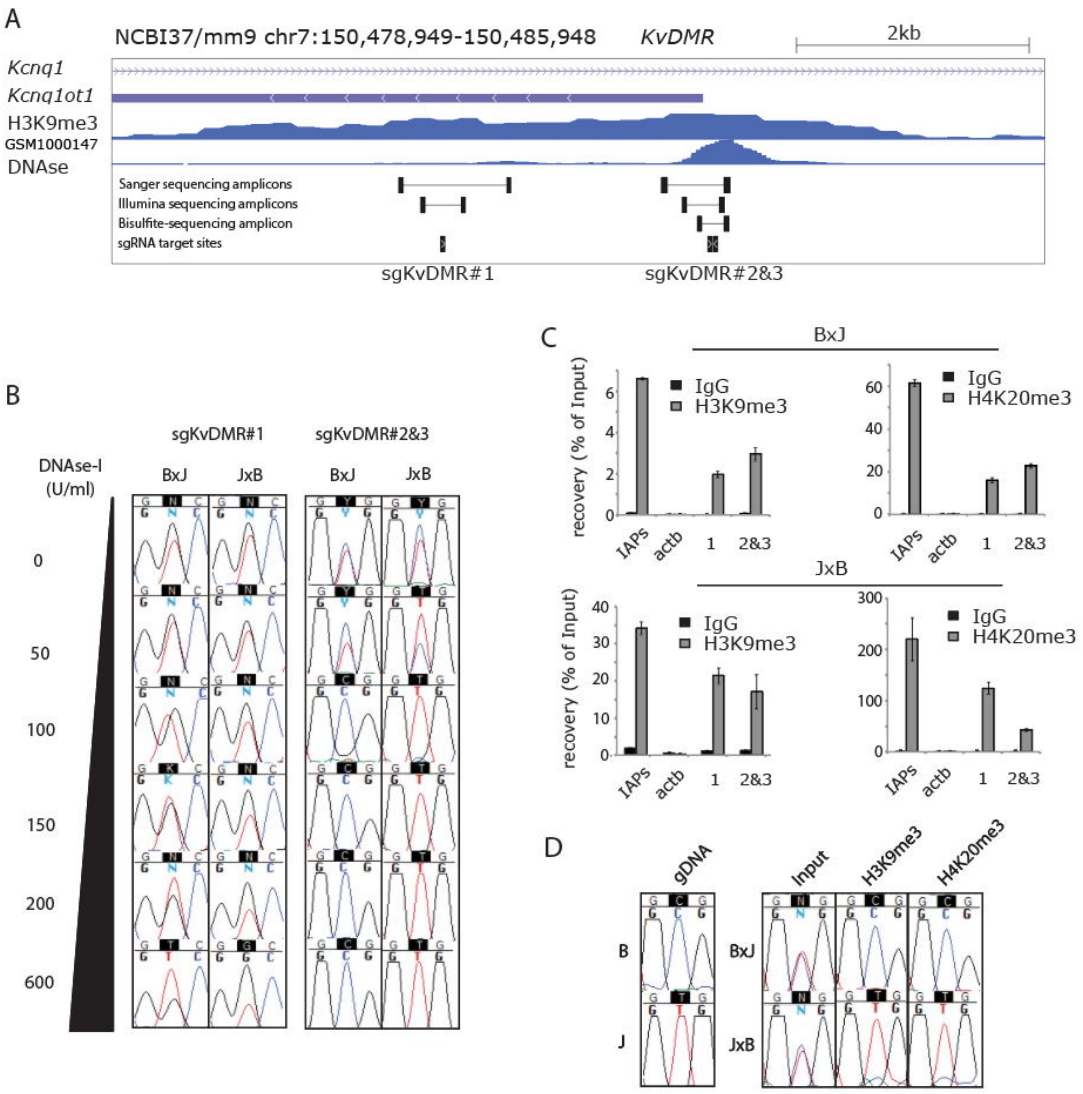


Figure 3.2 Allele-specific epigenetic characteristics at the maternally imprinted *KvDMR* locus. A: View of the *KvDMR* region including the sgRNA positions, Sanger and MiSeq amplicons. **B:** Allele-specific Sanger sequencing data after PCR amplification from digested nuclei. Primer locations for this Sanger PCR amplicon are indicated in panel A. **C:** Native histone ChIP data for H3K9me3 and H4K20me3 histone modifications for the *KvDMR* region. Error bars represent the Standard Deviation of three technical replicates. Intracisternal A particle (IAP) retrotransposons' qPCR primers were used as a positive control, *Actb* qPCR primers were used as a negative control. **D:** Allele-specific enrichment of H3K9me3 and H4K20me3 on the promoter of *Kcnq1ot1*. The gDNA feature indicates the specific SNPs associated with each parental mouse strain.

In addition, Gillian performed native histone ChIP for repressive histone modifications associated with imprinting. She demonstrated that there is an enrichment for H3K9me3 and H4K20me3 at the ICR and promoter within the *KvDMR* region, compared to the promoter of the mouse beta actin gene (*Actb*) as shown in **Figure 3.2.C**. PCR amplification and Sanger sequencing of the Input and IP material, revealed that the repressive histone modifications were enriched on the maternal allele (**Figure 3.2.D**). This assured us that imprinting on the *KvDMR* region was maintained in the two hybrid mESC lines we intended to use and defined the experimental setting for subsequent editing experiments.

The second selected locus was *Impact*, which is not found in a cluster, but is an individual maternally imprinted gene. *Impact* is one of the most highly expressed maternally-imprinted loci in the murine genome and is situated on chromosome 18 (Maupetit-Méhouas *et al.*, 2016), though its expression level has not been quantitatively assessed in the two hybrid mESC lines by myself. At the start of my PhD I designed an sgRNA that targets the promoter of *Impact* (sgImpact), in a site characterised by high sensitivity to DNase I. As for the KvDMR region, the position of sgImpact and the H3K9me3 ChIP and DNase I –seq ENCODE data from mESCs were obtained through the UCSC genome browser and shown in **Figure 3.3.A**.

As previously done for the *KvDMR* region, the same DNase I digestion assay was performed. The position of the corresponding Sanger sequencing amplicon is shown in **Figure 3.3.A**. The SNP from the paternal allele is preferentially digested and the SNP from maternal allele remains intact (**Figure 3.3.B**). This time the effect was not as profound as for the other two target sites from above, since 600 units of DNase I resulted in complete digestion in the BxJ line. In the JxB line traces of the paternal SNP remain even with 600 units of DNase I. This could be owed to partial loss of imprinting for this locus in the JxB line. Overall, we concluded that the paternal allele of the *Impact* locus is more accessible than the maternal one.

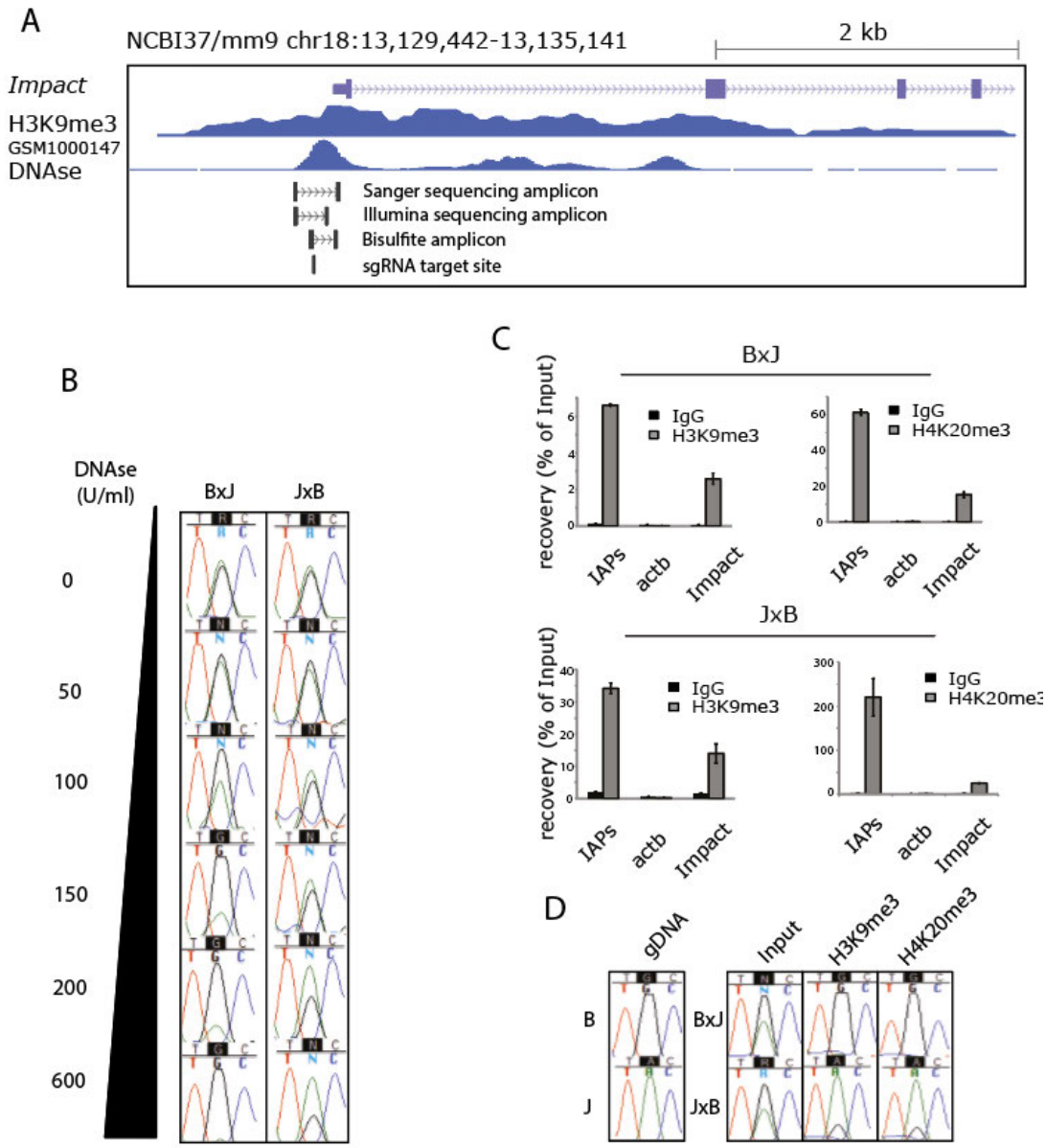


Figure 3.3 Allele-specific epigenetic characteristics at the maternally imprinted *Impact* locus. A: View of the *Impact* promoter and the position of the sgRNA target site, Sanger and MiSeq amplicons. **B:** Allele-specific Sanger sequencing data following PCR amplification from digested nuclei. Primer locations for this Sanger PCR amplicon are indicated in panel A. **C:** Native histone ChIP data for H3K9me3 and H4K20me3 histone modifications for the chosen target site. Enrichments were determined as for **Figure 3.2.C**. **D:** Allele-specific enrichment of H3K9me3 and H4K20me3 on the *Impact* promoter, as determined by Sanger sequencing of PCR amplicons from Input and ChIP materials. The gDNA feature indicates the specific SNPs associated with each parental mouse strain.

Native histone ChIP for H3K9me3 and H4K20me3 modifications showed that there is an enrichment for both histone modifications at the promoter of *Impact* compared to the *Actb* promoter, as shown in **Figure 3.3.C**. Following PCR amplification and Sanger sequencing of the Input and IP material amplicons, we demonstrated that H3K9me3 and H4K20me3 would only associate with the maternal allele (**Figure 3.3.D**). However, the DNase I profiles shown in **Figure 3.3.D** show that some of the paternal allele remains in heterochromatin-associated ChIP fractions from the JxB line. This raises the possibility of a partial loss of imprinting of *Impact* in this line only. Overall, it confirmed that the imprinting was maintained for this locus in the BxJ hybrid mESC line.

The third selected maternally-imprinted region in our study was the CpG island located upstream of *Inpp5f_v2*, which is a variant of the non-imprinted *Inpp5f* gene on chromosome 7 in mice characterised by differential allele-specific CG dinucleotide methylation (Choi *et al.*, 2005). Like *Impact*, *Inpp5f_v2* is not found in a cluster, but its level of expression is low, though it has not been quantitatively assessed in the two hybrid mESC lines by myself. Keerthi Chathoth designed sgInpp5fv2 within DNase I hypersensitive site, as shown in **Figure 3.4.A** along with data for H3K9me3 ChIP in mESCs from the ENCODE project in a similar manner as for the two previous imprinted targets.

In a similar manner to the other two loci, Gillian demonstrated that the SNP of the paternal allele is preferentially digested by increasing units of DNase I in **Figure 3.4.B**. In the BxJ line, traces of the paternal SNP can be detected even for 200 units of DNase I. This could also be due to partial loss of imprinting for *Inpp5f_v2* in the BxJ line, as discovered by subsequent findings (**Figure 3.9**). Thus, we were motivated to create an additional control to measure imprinting maintenance, which I will describe in detail in the next section. Since it was not possible to amplify from the amplicon material after digestion with 600 units of DNase I, the data is not available for this locus. However, there was an evident enrichment for H3K9me3 and H4K20me3 in the promoter of *Inpp5f_v2* (**Figure 3.4.C**).

In contrast to the two previously validated loci, the allele-specific association of the two repressive histone modifications was determined through restriction fragment length polymorphism (RFLP) analysis of ChIP amplicons performed by Gillian Taylor. **Figure 3.4.D** shows the digestion patterns of the two amplicons from the distinct parental genomes and input samples from each hybrid line (on the right hand side). To the left, the digested amplicons from IP samples from each histone ChIP are shown for BxJ and JxB. Every histone ChIP digestion pattern matches the pattern that corresponds to the maternal allele for either the BxJ or JxB line. Therefore, we concluded that the maternal allele is decorated with these repressive histones and the Inpp5f_v2 region was also maternally-imprinted in the hybrid mESC lines.

In the upcoming section I will describe the strategy that was used to target the three regions that were validated in this section. The experimental process begins with targeting experiments in the hybrid mESC lines and ends with quantification of allele-specific genome editing outcomes with the aid of a custom computational pipeline.

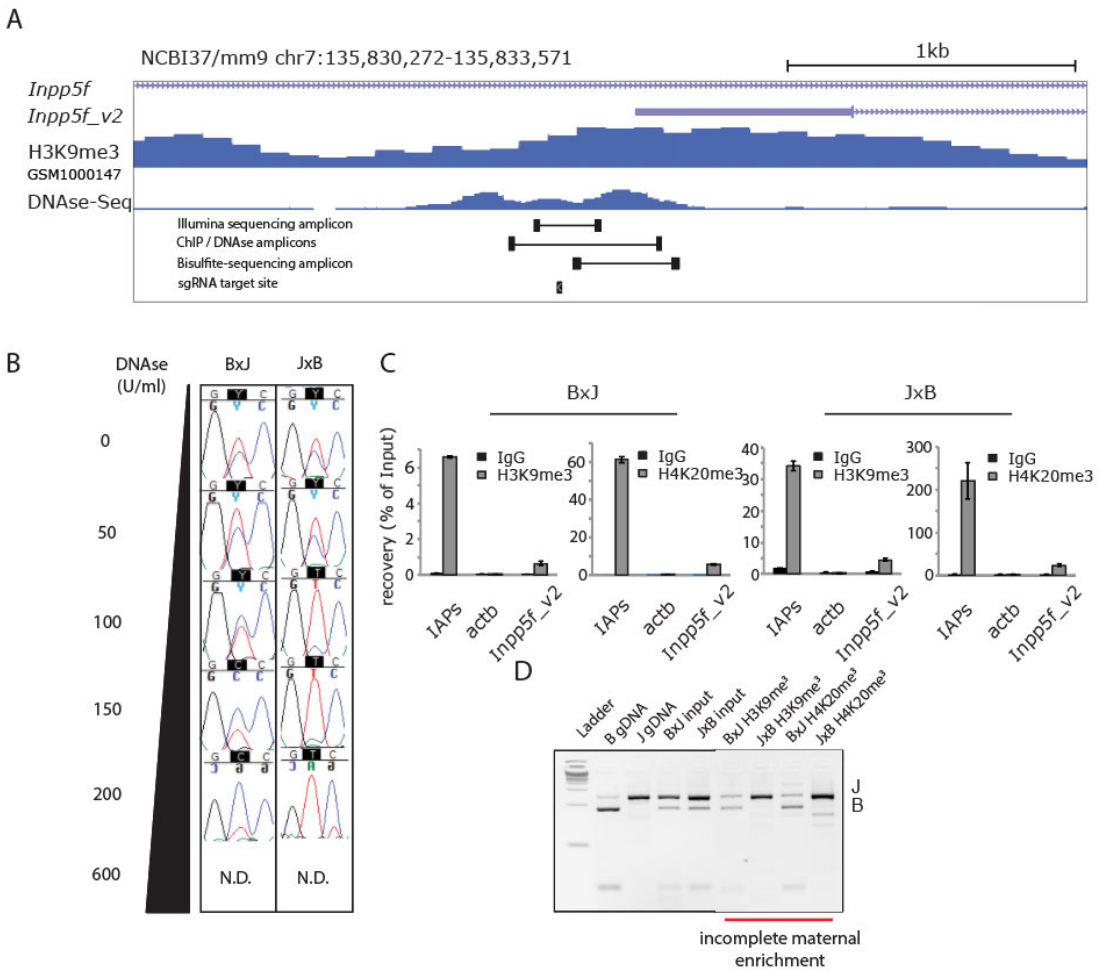


Figure 3.4 Allele-specific epigenetic characteristics at the maternally imprinted *Inpp5f_v2* locus. A: View of the *Inpp5f_v2* promoter and the position of the sgRNA target site, Sanger and MiSeq amplicons. **B:** Allele-specific Sanger sequencing data following PCR amplification from digested nuclei. Primer locations for this Sanger PCR amplicon are indicated in panel A. **C:** Native histone ChIP data for H3K9me3 and H4K20me3 histone modifications for the chosen target site. Enrichments were determined as for Figure 3.2.C. **D:** Allele-specific enrichment of H3K9me3 and H4K20me3 on the *Inpp5f_v2* promoter was determined by RFLP analysis of PCR amplicons generated from the Input and ChIP materials. The first initial of each hybrid line corresponds to the maternal mouse strain. The gDNA samples on the left hand site indicate the specific pattern associated with each parental mouse strain.

3.4 Setting up the allele-specific mutagenesis system

After validating the presence of imprinting on the 3 maternally-imprinted loci we intended to target, Keerthi Chathoth and I began performing targeting experiments in the reciprocal hybrid mESC lines.

I have been transfecting both hybrid lines in parallel with plasmid px459v2 (Addgene #62988) following cloning of the sgRNAs mentioned in the previous section. The px459v2 plasmid allowed the simultaneous expression of Cas9 and a sgRNA, along with Puromycin N-acetyltransferase (Pérez-González, Vara and Jiménez, 1983), which conferred Puromycin resistance to successfully transfected cells. In addition, I also included a carefully designed repair template per transfection mixture that would allow me to measure the efficiency of HDR in our system and compare its uptake frequency in the two different chromatin contexts. This was typically a single-stranded oligodeoxynucleotide (ssODN) containing a signature mutation that would abolish the PAM site upon incorporation and prevent consecutive rounds of cleavage by Cas9. Puromycin selection was performed to isolate successfully transfected cells, and genomic DNA samples were collected typically 4 days after transfection.

Following DNA quantification, amplicon sequencing libraries for the Illumina MiSeq platform were prepared for individual samples. This process would help us to accurately quantify genome editing outcomes in an allele-specific manner. After the sequencing reads were made available, we received invaluable help for the bioinformatic analysis from Alison Meynert of the Bioinformatics core of our Institute. Alison developed a custom computational pipeline to use the SNP information from the JF1 murine genome to determine the parental strain of origin for all reads. Subsequently, she detected the presence of the corresponding pre-engineered mutation from the relevant ssODNs I used from each target site or of small insertions and deletions (InDels) on the location of the cleavage site. These would be considered to be

HDR and NHEJ events, respectively. The total CRISPR/Cas9 mutagenesis frequency was calculated as the sum of the reads with evidence of HDR, NHEJ or both at the same time. A complete schematic of the experimental strategy and computational pipeline is provided in **Figure 3.5**.

Being aware that prolonged maintenance of mESCs in culture may lead to hypomethylation at a genome-wide level, I introduced an additional control to my targeting experiments. Imprints are specifically prone to erosion in female mESC lines, thus the BxJ and JxB lines we used were both male. However, when cultured in 2i conditions under simultaneous inhibition of MEK and GSK3, even male mESCs may become hypomethylated (Choi *et al.*, 2017). In the previous section we demonstrated that partial loss of imprinting might have occurred for *Impact* in the JxB line and for *Inpp5f_v2* in the BxJ line. Therefore, I developed a simple method to monitor DNA methylation near the target sites of interest.

Mock transfections with the px458 vector (Addgene # 48138) were performed in populations of both hybrid lines that were cultured in parallel to the cells used for targeting experiments with the CRISPR/Cas9 reagents. The imprinted target sites were not edited, due to the absence of a cloned protospacer in the sgRNA of the px458 vector. After the edited and mock-transfected cells were harvested, genomic DNA material from the mock-transfected cells would undergo bisulfite conversion with a commercially available kit. As a result, unmethylated cytosine residues would be changed to uracil residues whereas methylated cytosine residues would continue as such. A subsequent PCR amplification step converts uracil residues to thymines, which signal the presence of an unmethylated cytosine residue after sequencing of the end product (Frommer *et al.*, 1992).

Amplicon libraries for the Illumina MiSeq platform were prepared from the resulting converted DNA samples, in a similar manner to the non-converted samples from the targeting experiments. A full account for the differences in preparation of both types of Illumina amplicons is available in **Chapter 2.3**.

Once more, Alison Meynert helped me with the bioinformatic analysis of the generated MiSeq reads. She developed a special computational pipeline that would allow us to quantify the level of methylation across each template strand. After counting the number of methylated and unmethylated cytosine residues per read, Alison sorted the reads into three categories as shown in **Figure 3.6**. Therefore, we managed to record the numbers of unmethylated, partially and fully methylated reads to derive information about the DNA methylation levels, and consequently the degree of imprinting maintenance for each of our targeting experiments.

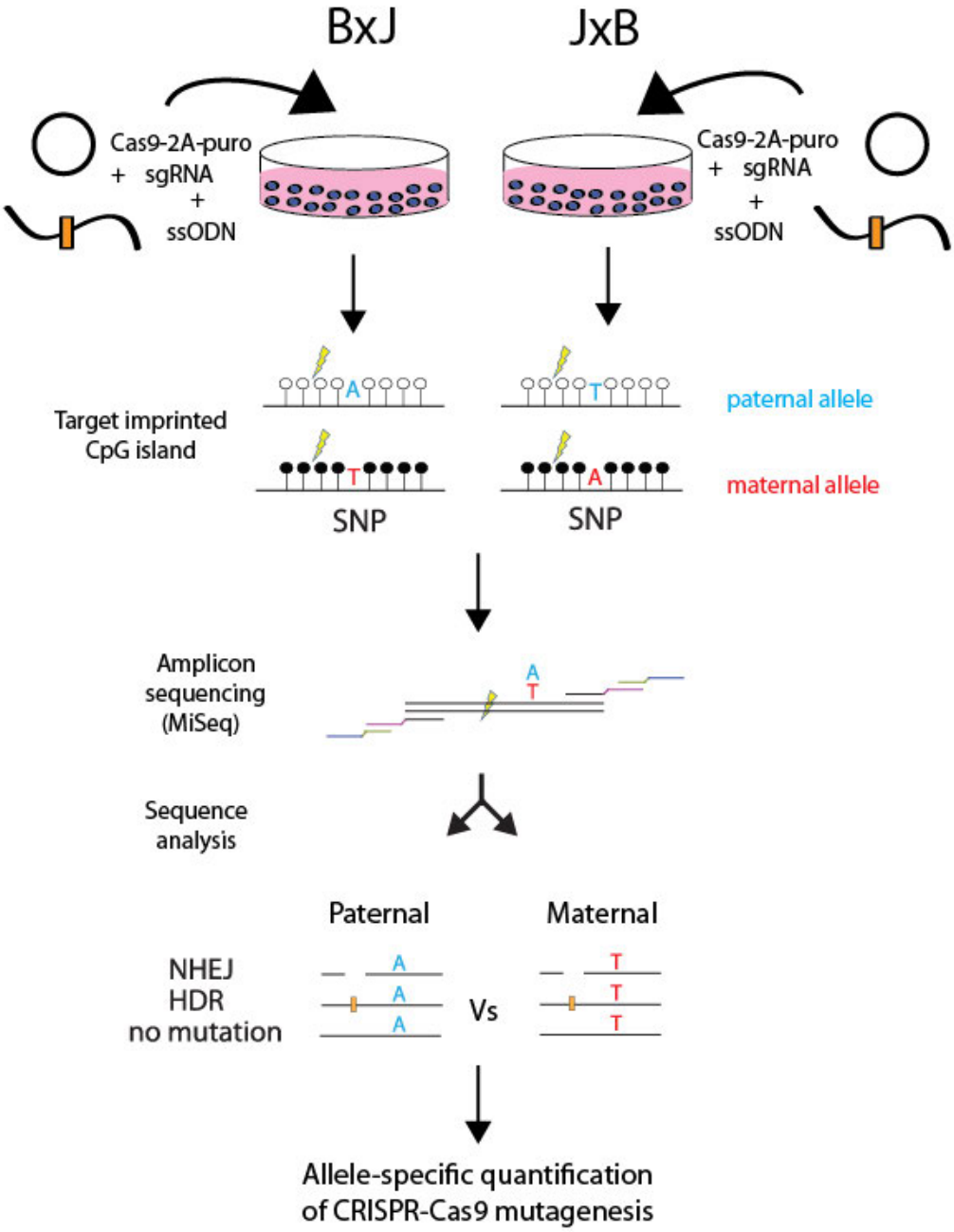


Figure 3.5 Schematic outlining the experimental strategy. Simultaneous targeting of maternally-imprinted loci was achieved by co-transfection of CRISPR plasmids and ssODNs. The pre-engineered special mutation is depicted in orange. The lightning bolt represents the Cas9 cut site. Illumina amplicons were generated by two consecutive rounds of PCR amplification to include adapter sequences (shown in purple) and barcode sequences (shown in blue). The target site for each sgRNA is conveniently positioned near a JF1 SNP within the differentially methylated target regions. MiSeq amplicons were always designed to cover both the SNPs and target sites.

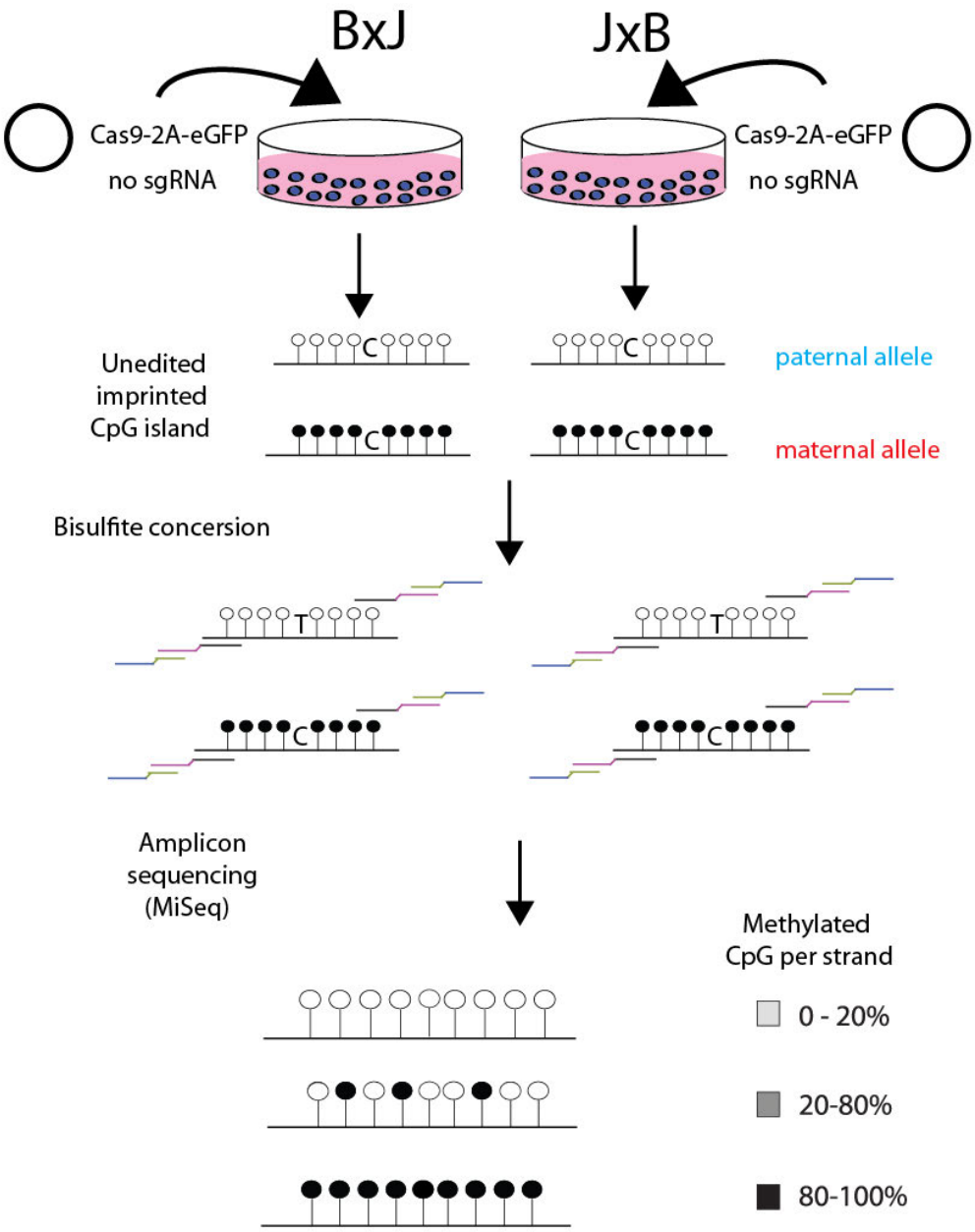


Figure 3.6 Schematic outlining the strategy for monitoring imprinting maintenance. Mock transfections without sgRNAs and ssODNs were performed and isolated gDNA material was subjected to bisulfite conversion. In theory, the maternal imprinted allele should be fully methylated while the paternal allele should be unmethylated. In practice though, the presence of partially methylated reads was also recorded. Similar to the targeting libraries, adapter and barcode sequences were assigned to bisulfite-converted amplicon libraries through a dual-round PCR procedure. The reads with 0-20%, 20-80% and 80-100% methylated CGs were classified as unmethylated, partially methylated and fully methylated reads, respectively.

3.5 Subtle chromatin-related effects after 4 days of targeting

After outlining our experimental strategy, I would like to introduce some of the early experimental results that were generated with the aforementioned methodologies. I began by transfecting both hybrid mESC lines in parallel, with the sgRNAs introduced in **Chapter 3.3**.

Being aware of the possibility that the imprinting modifications might be erased, as a consequence of maintaining the mESCs in culture (Choi *et al.*, 2017), all targeting experiments were conducted in hybrid mESCs that did not exceed passage number 12. The Puromycin selection process required 2 days, followed by an additional period of 2 days, which was necessary to allow the number of successfully transfected cells to increase. This was crucial for harvesting sufficient genomic DNA material for downstream preparation of amplicon sequencing libraries.

By analysing the corresponding MiSeq data, we were able to make the following observations. Using the SNPs to distinguish maternal from paternal reads, we calculated mutation frequencies separately for maternal versus paternal alleles. We then expressed them as a ratio of paternal / maternal to determine the allele-specific bias. Firstly, it was evident that the non-silenced paternal (shown in blue) allele was gaining mutations by CRISPR/Cas9 at a higher frequency than the inaccessible maternal allele (shown in red). **Figure 3.7** indicates that all 3 sgRNAs destined for the *KvDMR* region yielded significant differences in editing efficiencies between the two alleles, most notably in the BxJ hybrid line. This could be owed to the fact that imprinting was maintained at a higher degree for the *KvDMR* region in the BxJ line than in the JxB line (**Figure 3.9.A**). However, the observed allelic bias was quite subtle.

Secondly, this subtle effect was not always observed for these long targeting experiments, as pointed out in **Figure 3.8.A**. We were initially

surprised to discover a lack of allele-specific editing preferences on the promoter region of *Impact*. Upon closer investigation of the MiSeq bisulfite amplicon sequencing data we noted a large drop in sequencing reads with fully methylated CG dinucleotides and concluded that this target locus had undergone partial loss of imprinting (**Figure 3.9.B**). This could account for the lack of observed allelic bias in this case and explain the observed loss of allele-specific accessibility observed in **Figure 3.3.B**.

Interrogation of the DNA methylation levels on the promoter of *Inpp5f_v2* revealed a larger hypomethylation on this locus (**Figure 3.9.C**) compared to *Impact*. The level of loss is greater for the BxJ line this time, and may provide an explanation for the persistence of the paternal allele on this line in the DNase I assay in BxJ shown in **Figure 3.4.B**. The targeting data for *Inpp5f_v2* is therefore not shown.

Finally, we also revealed the absence of allele-specific bias in editing efficiency for a non-imprinted target locus. We selected to target the N-terminus of *Ncaph*, which is not known to undergo imprinting and for which a previously validated sgRNA was available. In **Figure 3.8.B** we demonstrate that in contrast to the maternally-imprinted target sites, there is no allele-specific editing preference for *Ncaph* in both hybrid mESC lines.

To further confirm that epigenetic chromatin modifications are solely responsible for the observed allelic bias, we attempted to remove DNA methylation at a global level in the two hybrid mESC lines. As I will describe in the following section, we were keen to induce loss of imprinting at our selected target sites and observe if this would affect the allelic bias in our allele-specific genome editing monitoring system.

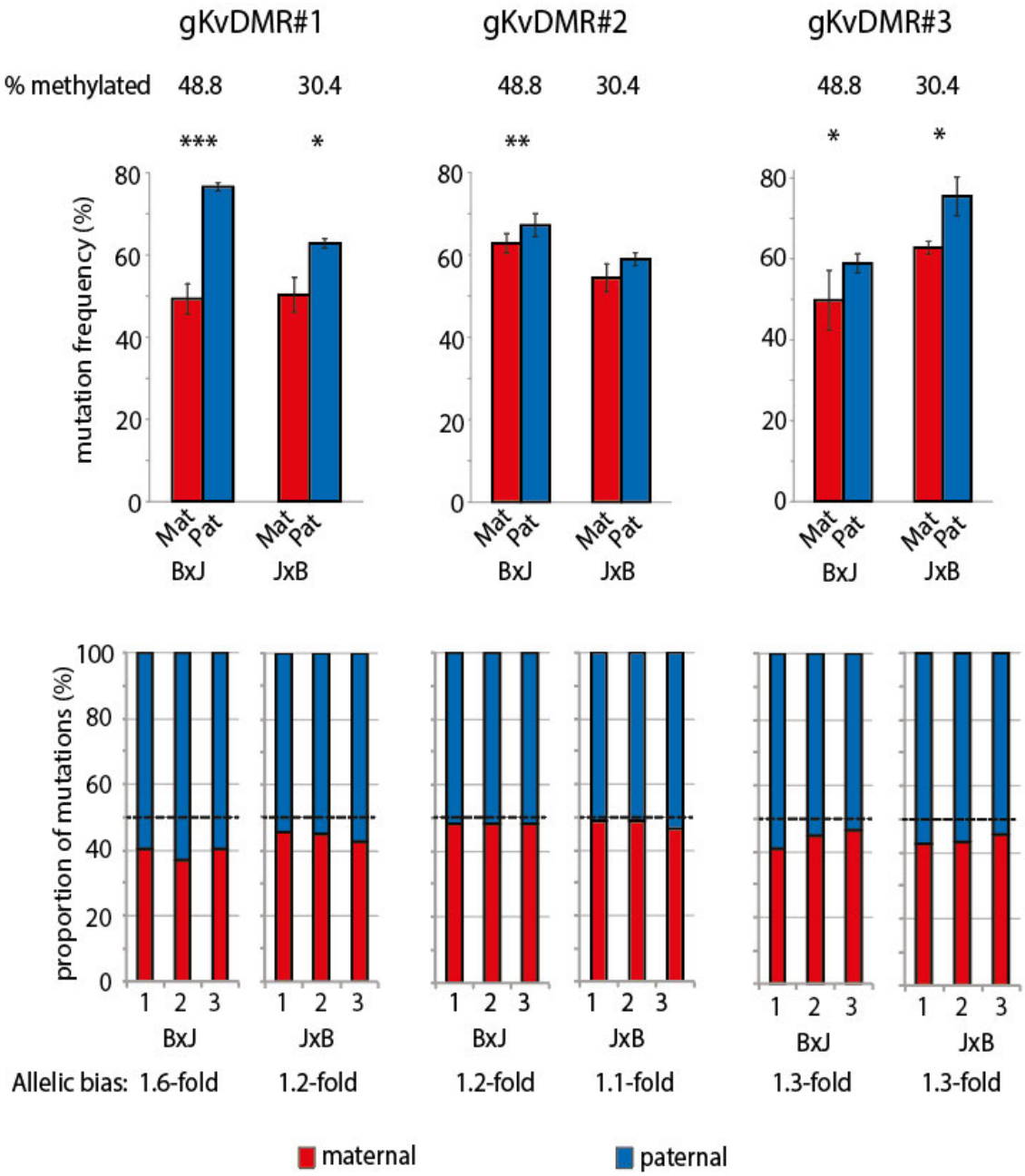


Figure 3.7 Higher editing frequencies on the accessible paternal allele after a 4 day-long exposure to CRISPR/Cas9. A: Allele-specific editing frequencies for sgKvDMR#1, sgKvDMR#2 and sgKvDMR#3. Error bars represent the Standard Deviation of three experimental replicates. Asterisks correspond to p-values derived from paired two sample t tests for the difference in editing efficiencies on the maternal and paternal allele. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. The percentage of reads with fully methylated CG dinucleotides is shown above every histogram to show the degree of imprinting maintenance during each targeting experiment.

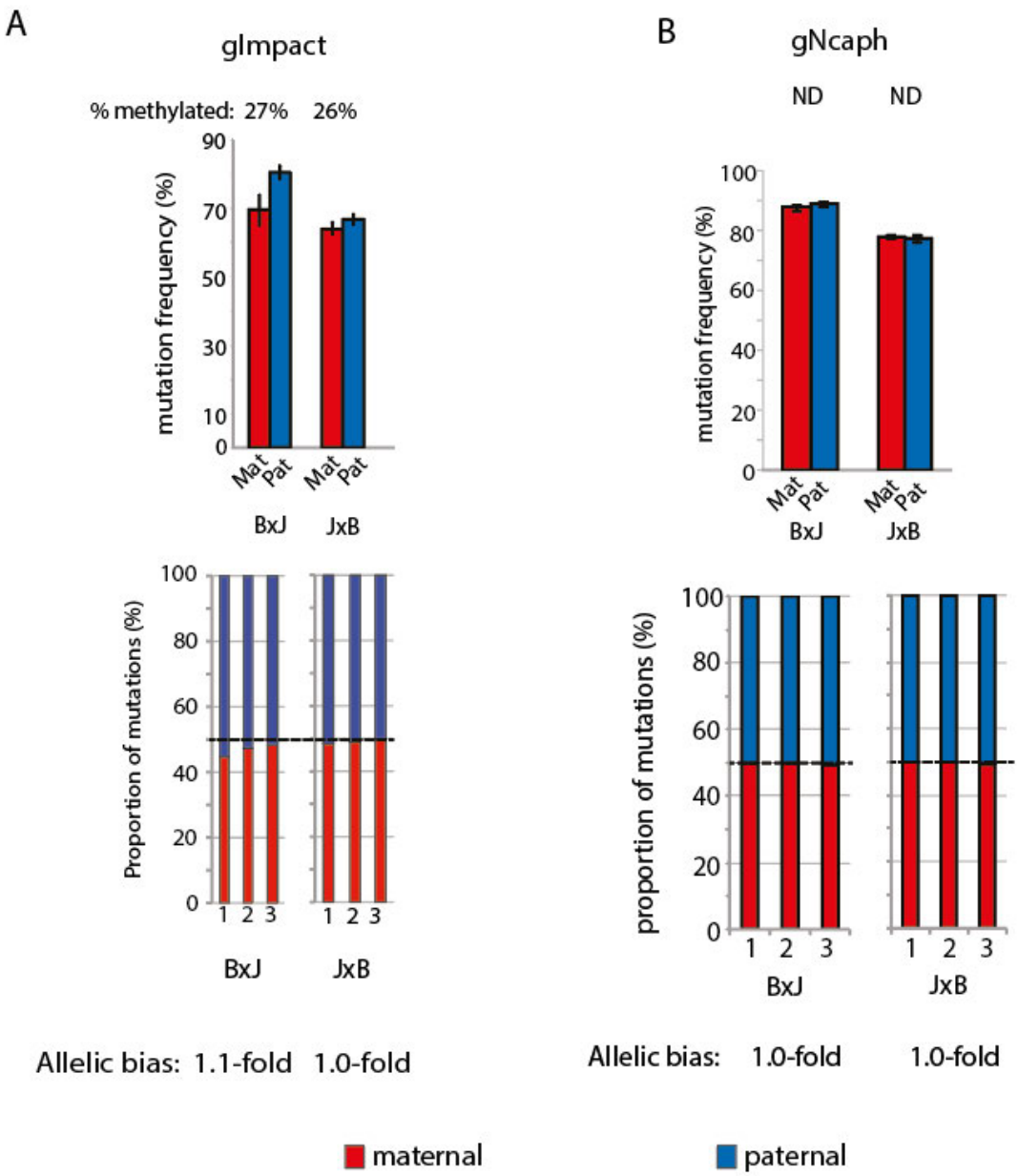


Figure 3.8 Loss of imprinting results in lower allelic bias and resembles a non-imprinted target site. A: Allele-specific editing frequencies for sgImpact. Error bars represent the Standard Deviation of three experimental replicates. Paired two sample t tests for the difference in editing efficiencies on the maternal and paternal allele showed the absence of a statistically significant difference. The percentage of reads with fully methylated CG dinucleotides is shown above every histogram to show the degree of imprinting maintenance during each targeting experiment. **B:** Allele-specific editing frequencies for the sgRNA at the N-terminus region of *Ncaph*. Same as for sgImpact.

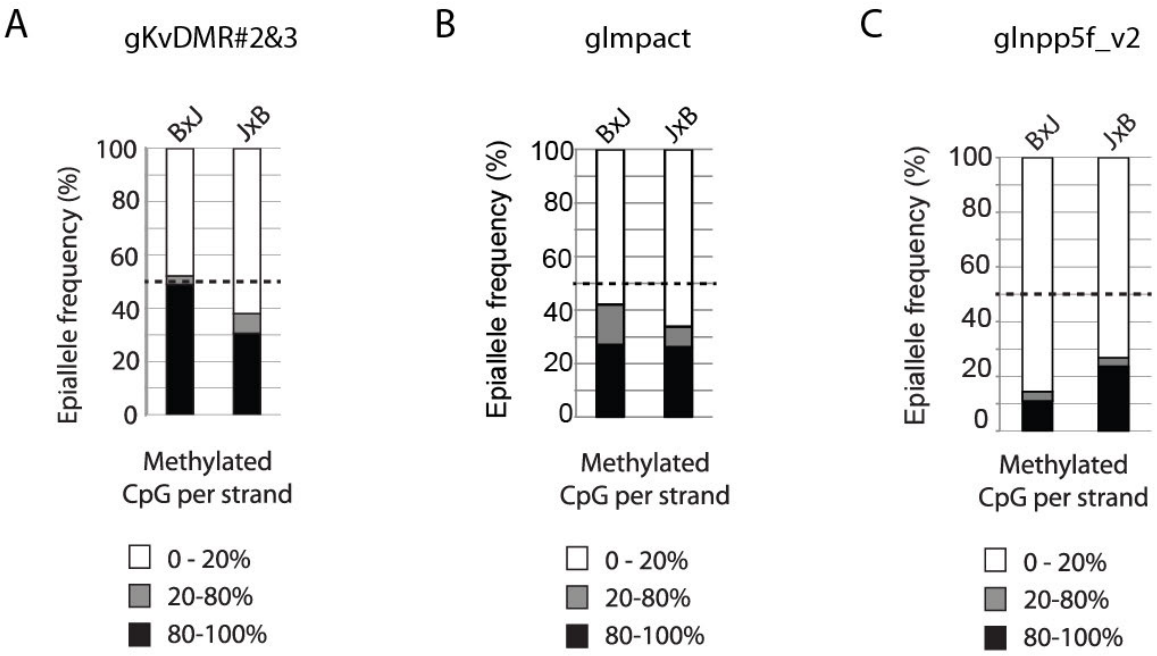


Figure 3.9 Imprinting is maintained at different degrees across different target loci. A: CG dinucleotide methylation levels at the *KvDMR* region in both hybrid mESC lines after a mock transfection with px458. This locus had the greatest level of imprinting retention, especially in the BxJ line. The amplicon contains 10 CG dinucleotides. **B:** CG dinucleotide methylation levels at the promoter region of *Impact* in both hybrid mESC lines after a mock transfection with px458. A partial loss of imprinting is evident in both hybrid lines. The amplicon contains 7 CG dinucleotides. **C:** CG dinucleotide methylation levels at the promoter of *Inpp5f_v2* in both hybrid mESC lines after a mock transfection with px458. With very few methylated reads obtained from BxJ, this is the region where we discovered the most detrimental degree of loss of DNA methylation and imprinting. The amplicon contains 22 CG dinucleotides.

3.6 Epigenetic modifications influence the levels of allelic bias

Following the observation of a subtle yet reproducible preference for editing on the accessible, euchromatic paternal allele, it was necessary to demonstrate that this preference was owed to the allele-specific epigenetic modifications.

To achieve this, I tried to induce loss of imprinting by culturing mESCs in 2i media supplemented with vitamin C. The aforementioned culture conditions have been reported to induce a global loss of DNA methylation in cultured E14 mESCs after a continuous treatment of 6 days by promoting the activity of Tet enzymes (Blaschke *et al.*, 2013; Walter *et al.*, 2016). I began with a low passage BxJ hybrid line, cultured it for several days in the standard 1i culture medium to increase the passage number, and then I switched the culture medium to 2i. I continued increasing the passage number under these culture conditions to begin removing some of the DNA methylation (Choi *et al.*, 2017). Once passage 20 was reached, I also added an appropriate concentration of vitamin C in the 2i culture medium and continued culturing the cells for an additional period of 6 days. Thus, I generated a high passage BxJ cell line that would be hypomethylated as a consequence of the 2i/vitamin C culture treatment.

As shown in **Figure 3.10.A**, I also targeted the same BxJ hybrid line, while it was at a low passage and had not been treated like the high passage BxJ hybrid line. Interrogation of the Illumina bisulfite amplicon sequencing data revealed a trend towards loss of methylation at all 3 loci, but to very different degrees. The reason for this locus specificity is not clear. **Figure 3.10.B** shows that DNA methylation was largely maintained in the *KvDMR* region, where the allelic bias in the CRISPR mutagenesis was also maintained. Similarly, the allelic bias was also present on *Impact*, where only partial loss of DNA methylation was detected. Finally, DNA methylation was almost completely

erased from the *Inpp5f_v2* promoter region. This resulted in equal frequencies of CRISPR mutagenesis between the two alleles of *Inpp5f_v2*, and the loss of allelic bias (**Figure 3.11**).

Thus, we concluded that the previously observed allelic bias was owed to the allele-specific epigenetic modifications that can be found in the targeted maternally-imprinted loci. The effect is also reversible, since allelic bias can be eliminated through the removal of DNA methylation near the target site.

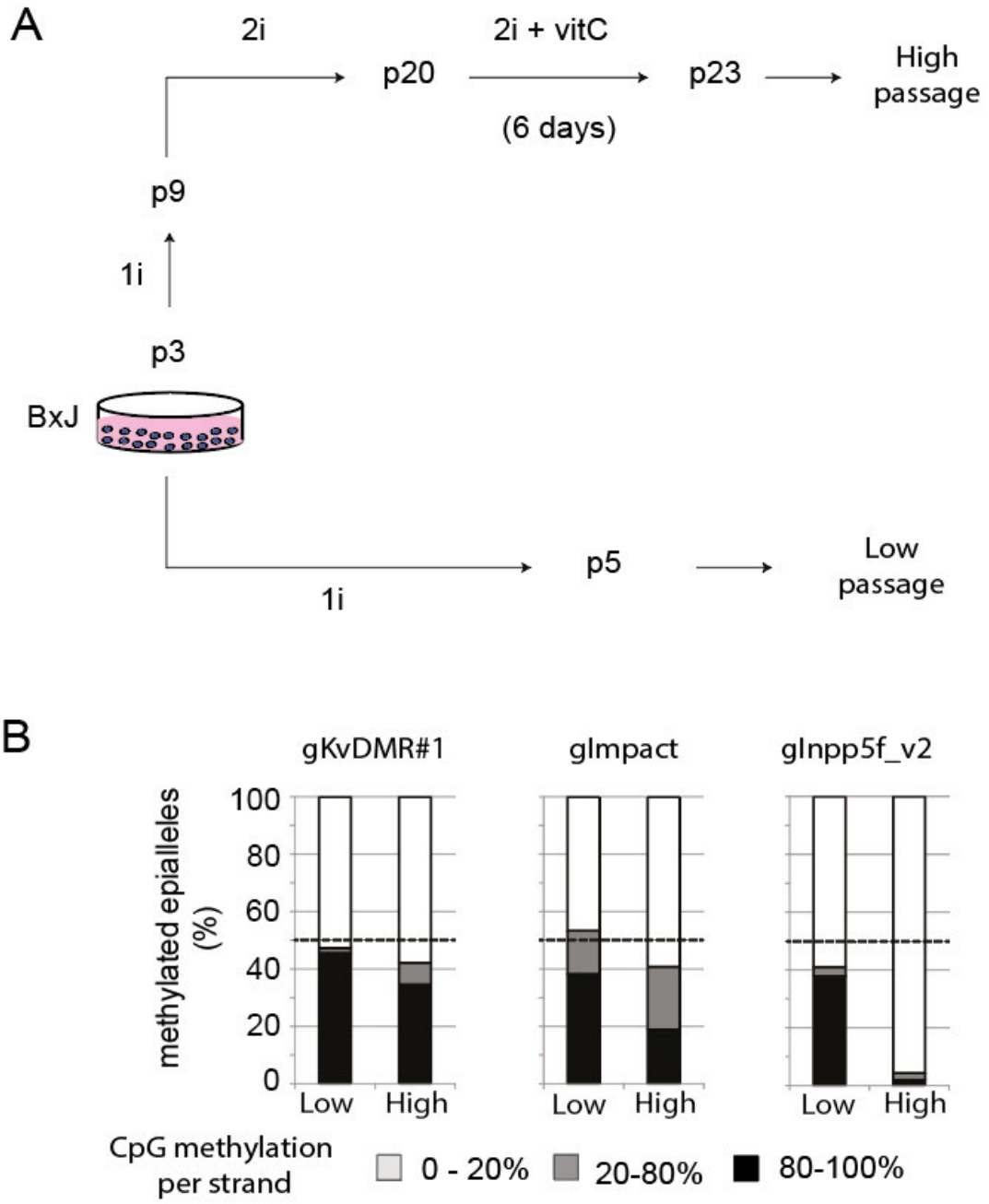


Figure 3.10 Inducing loss of imprinting in BxJ mESCs. A: Outline of the DNA hypomethylation treatment that was used to generate the high passage BxJ line. The low passage line was untreated and both lines were targeted in parallel. **B:** CG dinucleotide methylation levels at three target loci in the high and low passage BxJ lines after a mock transfection with px458. There are less methylated CGs in the high passage line compared to the low passage line, but the greatest loss is observed in the *Inppf5_v2***.**

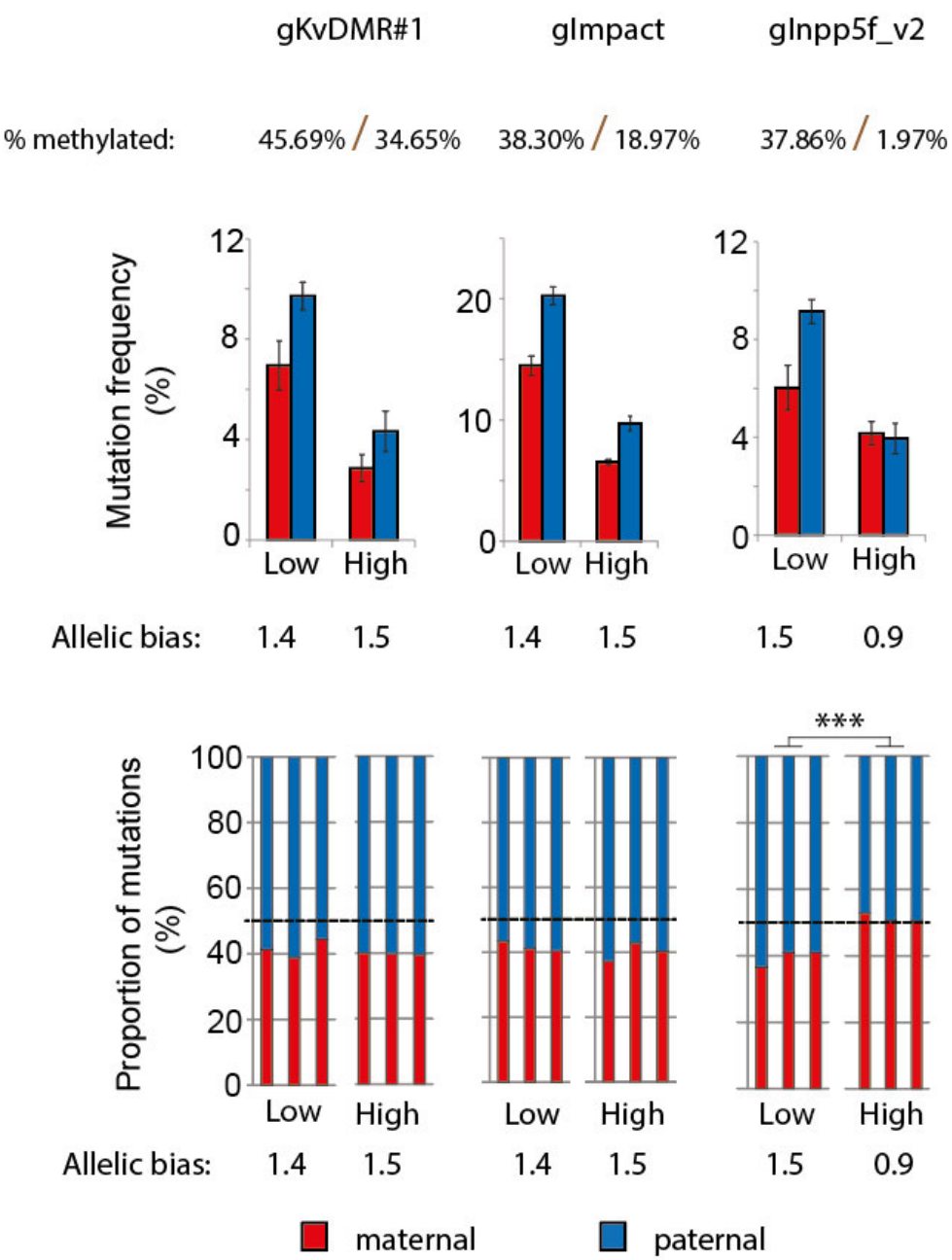


Figure 3.11 Allele-specific editing bias is owed to allele-specific epigenetic modifications. Allele-specific editing frequencies for sgKvDMR#1, sgImpact and sgInpp5f_v2 in both BxJ hybrid lines. Error bars represent the Standard Deviation of three experimental replicates. Asterisks correspond to p-values derived from unpaired t tests on the fold-difference between mutation frequencies on maternal and paternal alleles in low- compared to high-passage cells. ***p < 0.001. Due to the absence of Puromycin selection the editing efficiency appears lower than for **Figures 3.7&8.** Maintenance of the high passage BxJ line in 2i culture conditions resulted in a smaller transfection efficiency that has further reduced the overall mutation frequency**.**

3.7 Discussion

Despite the fact that genome editing with CRISPR/Cas9 is nowadays used to introduce genetic modifications in a plethora of organisms ranging from plant (Char *et al.*, 2017) to mammalian species (Nelson *et al.*, 2019), little is known about how eukaryotic chromatin compaction affect the most widely-used genome editing tool.

In this chapter I have explained how we successfully repurposed a naturally occurring epigenetic mechanism into a system that would allow us to assess the effect of chromatin modifications on targeted mutagenesis. In the past, genomic imprinting has provided insights into numerous epigenetic phenomena, including the function of long non-coding RNAs in chromatin modification (Filson *et al.*, 1993; Wutz *et al.*, 1997, 2001) and the role of CTCF binding as an insulator element (Bell and Felsenfeld, 2000; Hark *et al.*, 2000; Yusufzai *et al.*, 2004). Our study represents the first use of this system, to our knowledge, to study the effects of different chromatin states on mutagenic processes. It is therefore evident that it is a very attractive, natural and timeless system that may play a key role in other great discoveries in the future.

Prior to the commencement of my PhD project other research groups have published evidence that chromatin impedes mutagenesis by CRISPR/Cas9. The position of the PAM site relative to individual nucleosomes seems to be the factor that determines if the Cas9 endonuclease will generate a double-strand break (Hinz, Laughery and Wyrick, 2015; Horlbeck *et al*., 2016; Isaac *et al*., 2016). Nevertheless, these findings were established from *in vitro* studies assessing one nucleosome at a time. It is of paramount importance to investigate what is happening in living mammalian cells, where the genome is folded into higher order structures in three dimensional space.

Notably, evidence for the inhibitory effect of chromatin compaction on CRISPR/Cas9 mutagenesis have also emerged from studies that used cell-based systems (Chen *et al*., 2016; Daer *et al*., 2017). However, both of these

studies relied on exogenous reporter constructs and chemical induction for compaction and de-compaction of the wider regions where the reporters were situated. As a consequence, euchromatic and heterochromatic conditions were interrogated in distinct populations of cells, which creates a big caveat.

In contrast to the subtle results we obtained from our initial prolonged targeting experiments, where allelic bias would not exceed 1.6-fold, Chen *et al.* (2016) reported that the magnitude of effect they observed was up to 2.7-fold. CRISPR/Cas9 was also found to associate and edit the euchromatic state of the luciferase construct that was employed by Daer *et al.* (2017) by up to 5-fold. One possible explanation for this inconsistency is that our experimental system allows editing of both open and closed chromatin regions in the same cell population, instead of assessing the same sequence before and after perturbing chromatin in separate cell populations. This quality of our system abolishes any type of bias from transfecting separate cell populations, such as differences in the uptake efficiency of the introduced reagents. It also provides a convenient internal control that eliminates potentially confounding factors such as cell cycle stage and availability of trans-regulator molecules.

The caveat of our system is the attrition of the imprinting marks from the selected maternally-imprinted loci. We have already witnessed this effect in different degrees among different target loci, despite performing our experiments in cells that never exceeded passage 12 in culture. To minimise this as much as possible, we used male BxJ and JxB reciprocal hybrid mESC lines that we cultured in the absence of serum under inhibition of GSK3 alone. We also developed a method to monitor the degree of imprinting maintenance in every individual targeting experiment, by interrogating the DNA methylation status of a certain number of CG dinucleotides in close proximity to the target site. This assured us that we can reliably report on the effect that different epigenetic states exert on CRISPR mutagenesis *in vivo*, in a sensitive and quantitative manner.

A similar system has recently emerged, where a number of endogenous target sites were selected in human HeLa and HEK293 cells (Kim and Kim,

2018). These random sites were selected after extensive analysis of publicly available DNase Hypersensitivity data from the ENCODE project. Their key feature was that they were present in two copies in the human genome, one of which was accessible and the other inaccessible. Contrary to what has been reported so far, Kim and Kim (2018) argued that chromatin does not affect genome editing with CRISPR/Cas9 when the sgRNA is perfectly complementary to the assessed target site. It is worth mentioning though that the timeframe of those targeting experiments was not clearly specified. In addition, the wider chromatin context where the selected target sequences were embedded was not reported, or the extent of the regions that were accessible/inaccessible for each copy of the different target sites they picked.

All of the above inspired us to continue investigating and work towards gaining mechanistic insights of the inhibitory effect we observed. Daer *et al* (2017) suggested that the degree of inhibition may be regulated by additional factors that were not assessed at the time. Thus, we decided to use our system to discover what other factors may affect the way chromatin modifications regulate CRISPR mutagenesis. In the next chapter I will investigate how the time of exposure to the genome editing reagents and the concentration of the Cas9 endonuclease may lead to higher allelic bias ratios and a higher editing frequency of the paternal euchromatic allele in our system.

Chapter 4

Study of the additional parameters that influence the effect of chromatin on CRISPR/Cas9 mutagenesis

4.1 Introduction

The effects we have observed during a 4-day exposure to CRISPR/Cas9 (**Figure 3.7** and **Figure3.8**) were subtle. As mentioned earlier in **Chapter 1.3.3**, it is possible that certain experimental conditions may influence the way chromatin modifications affect the action of CRISPR/Cas9. For example, Knight *et al.* have demonstrated that the kinetics of Cas9 diffusion are slower within heterochromatic regions (Knight *et al*., 2015). Our previous targeting experiments focused on a single time point: 4 days after lipofection. Therefore, we decided to quantify and compare CRISPR/Cas9 activity at maternally-imprinted regions for shorter periods of exposure. Another important factor to consider is the concentration of the Cas9 endonuclease, which may have been different between the separate editing experiments described in the previous chapter.

In addition, evidence from a more recent *in vitro* study by Hinz, Laughery and Wyrick (2016) suggested that nucleosome positioning is key to reducing off-target site mutagenesis by CRISPR/Cas9. By changing individual nucleotides in the protospacer region, and thus reducing complementarity between the sgRNA molecule and the target site, the authors simulated off-target recognition using naked DNA and nucleosome substrates. It was established that single nucleotide PAM-proximal mismatches were only reducing Cas9 activity on nucleosome substrates. This inspired us to investigate the effect of chromatin modifications on off-target mutagenesis by making use of the system we previously established. We also decided to examine different combinations of more than one mismatches in the protospacer. By determining the key players for achieving target-specific editing at high frequency, we aim to improve our understanding of the mechanism of action of CRISPR/Cas9 within mammalian cells and contribute towards promoting the use of genome editing in the clinic.

The work presented in this chapter was primarily my own, except for the following notable contributions. Elizabeth Freyer helped me with flow cytometry. Gillian Taylor contributed by cloning the plasmid constructs for the dCas9 ChIP. Andrew Wood harvested some critical time points in the middle of the night during the initial *Impact* promoter and dCas9 ChIP time course experiments. As mentioned before, Alison Meynert conducted the bioinformatics analysis for the NGS data.

4.2 Brief exposure to CRISPR/Cas9 elevates allelic bias

To test if chromatin modifications delay CRISPR mutagenesis in our previously described and validated system based on genomic imprinting, I performed a time course experiment in the BxJ hybrid mESC line. Initially, I targeted the promoter region of *Impact* because this was the locus where we previously observed the lowest allelic bias following 4 days of exposure to CRISPR/Cas9 (**Figure 3.8.A**).

Selection of the precise time points for this experiment was of paramount importance. In a former study Kim *et al.* (2014) compared the kinetics of mutation frequency for two different types of genome editing reagent delivery in K562 cells. Namely, the reagents were Ribonuclear Particles (RNPs) and plasmids. It was established that the use of plasmid DNA resulted in a slower accumulation of InDels than purified Cas9 protein. As shown by Kim *et al.,* 2014, when Cas9 is expressed from a plasmid the expressed protein appears at 6 hours after transfection and continues to accumulate for up to 24 hours (Kim *et al.,* 2014). At 48 hours after transfection the level of Cas9 protein is reduced. We were therefore interested to quantify allele-specific mutagenesis during this period.

As shown in **Figure 4.1.A**, six experimental sets of triplicates were simultaneously transfected with sgImpact in the px459v2 puromycin vector. Each set was harvested at the indicated time points after transfection. In contrast to the experiments described in the previous chapter, Puromycin selection was not performed in this case, due to the absence of enough time to complete the selection process. Consequently, the overall mutagenesis frequency was reduced due to the higher frequency of reads originating from untransfected cells.

Upon averaging the mutagenesis information for the two assessed alleles we could clearly see that there was an increase in mutagenesis frequency. This is shown in **Figure 4.1.C** in the right-hand side y-axis of the graph. Allele-specific quantification of mutations revealed higher editing frequencies on the paternal allele throughout the assessed period of time (**Figure 4.1.B**). However, we observed a trend where mutagenesis on the maternal allele would also increase steadily during the time course experiment. This demonstrates that repressive chromatin modifications do not completely abolish CRISPR/Cas9 mutagenesis. The data plotted on the right-hand axis of the graph in **Figure 4.1.C** suggest that the allelic bias of CRISPR/Cas9 mutagenesis decreases with time. One-way Analysis of Variance (ANOVA) followed by Tukey's HSD test were performed to assess if there are significant differences in allelic bias from individual replicates between time points. The only statistically significant difference that emerged was between 16 hours and 48 hours after transfection. What was also striking from **Figure 4.1.C** was that the highest allelic bias was observed at 16 hours after transfection. This inspired me to inspect if this phenomenon is also true for more genomic sites. As for the promoter region of *Impact*, I used sgKvDMR#1 and sgInpp5f_v2 to target two other maternally imprinted regions. Separate sets of triplicates were harvested at 16 hours and 4 days after lipofection in the absence of puromycin selection. The results are shown in **Figure 4.2.A** and **Figure 4.2.B** for sgKvDMR#1 and sgInpp5f_v2, respectively.

Consistent with our previous findings for sgImpact, at 16 hours after transfection the paternal non-imprinted gene was edited at a higher frequency

than the maternal allele. The magnitude of effect was larger at 16 hours after transfection and was greatly reduced at 4 days after transfection, as shown in **Figure 4.2.C** for sgKvDMR#1. There was one occasion where the difference observed between the 16-hour and 4-day timepoint did not reach statistical significance, in the condition with sgInpp5f_v2 in the BxJ line (**Figure 4.2.D**). This could be owed to the drastic reduction in CpG methylation that we also observed for the promoter region of *Inpp5f_v2* in the BxJ line that was used for this experiment, as shown on the top right-hand side of **Figure 4.2.B**.

Overall, the data from three independent target loci suggest that allelic bias is higher for short periods following transfection. We therefore concluded that the frequency of CRISPR/Cas9 mutagenesis is impeded in less accessible chromatin regions in a time-dependent manner. If expression of Cas9 from a plasmid vector requires time for Cas9 protein levels to increase within the cell, it is possible that the concentration of Cas9 may have a contribution to the accentuation of this phenomenon. This will be investigated in the following section.

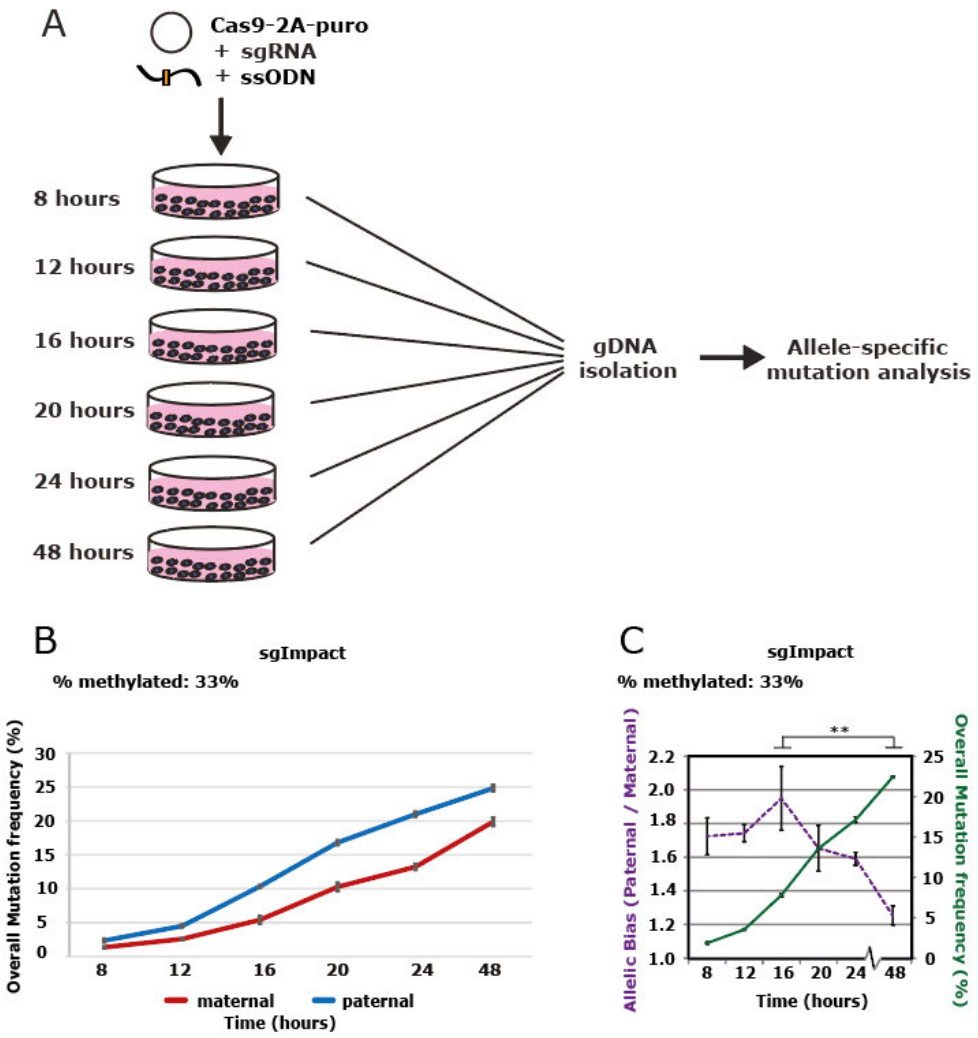


Figure 4.1 Chromatin modifications impair the kinetics of CRISPR/Cas9 mutagenesis on the promoter of *Impact*. A: Schematic of the experimental time frame for the time course experiment. BxJ cells were transfected px459v2 containing sgImpact and a ssODN. Cells were harvested at the indicated 4-hour intervals and genomic DNA was isolated for targeted amplicon NGS and allele-specific quantification of Cas9 mutagenesis. **B:** Allele-specific mutagenesis frequencies plotted against time. The paternal accessible allele is shown in blue and the maternal inaccessible allele is shown in red. Editing is more frequent on the paternal allele, but both alleles demonstrate a similar trend of increase in mutation frequency over time. The percentage fully methylated reads shown above implies the degree of imprinting maintenance during this experiment. **C:** Comparison of the overall mutation frequency and allelic bias. While mutagenesis (green) increases with time, allelic bias (purple) decreases. Asterisks for Tukey's HSD test p-values per pairwise comparisons **p < 0.01.

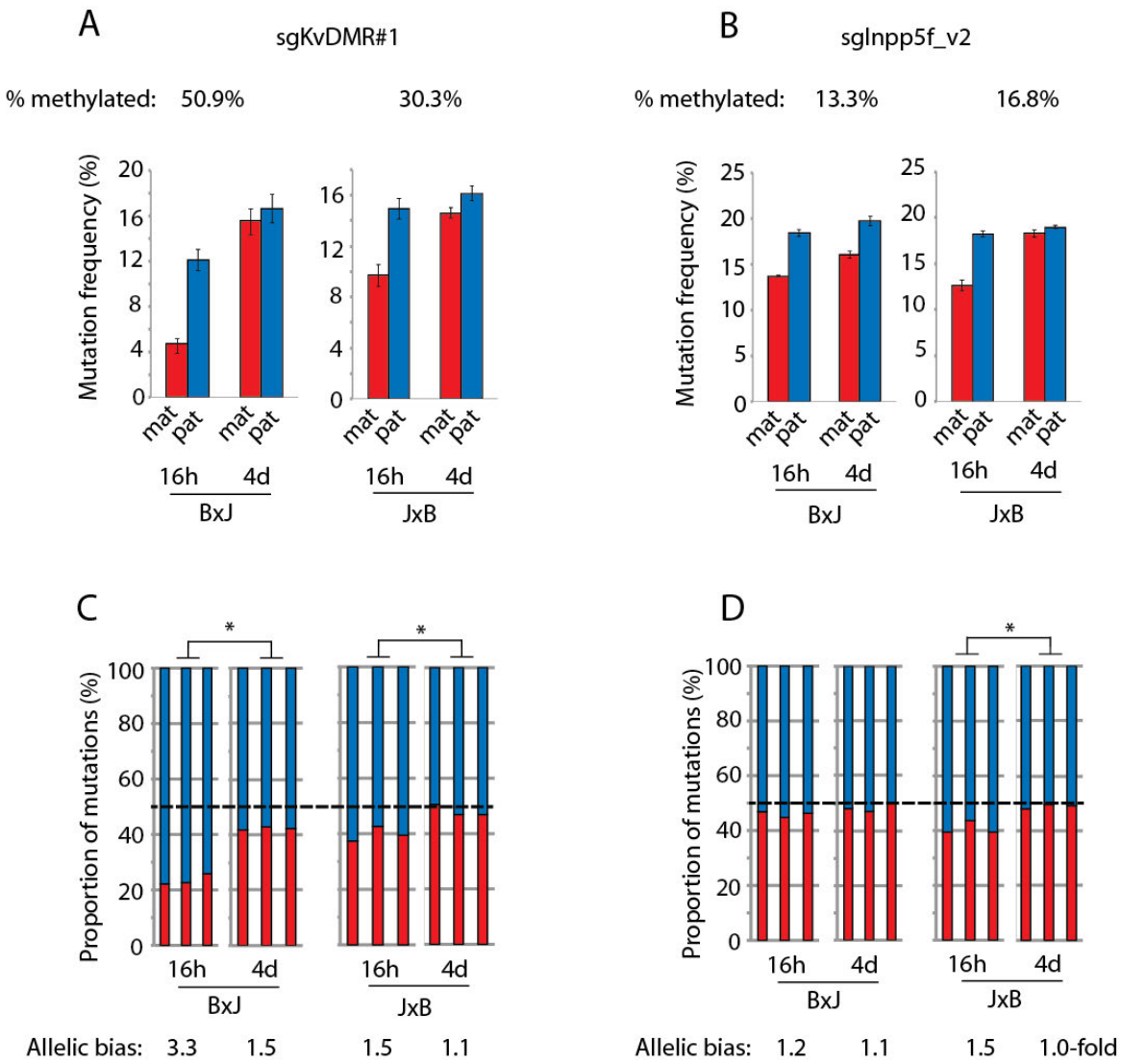


Figure 4.2 Chromatin modifications also impair the kinetics of CRISPR/Cas9 mutagenesis on *KvDMR* and *Inpp5f_v2*. A: Allele-specific editing frequencies for sgKvDMR#1 for 16 hours and 4 days after lipofection. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above every histogram to show the degree of imprinting maintenance. All the experiments in this figure were performed in the absence of selection for successfully transfected cells. **B:** Allele-specific editing frequencies for sgInpp5f_v2 for 16 hours and 4 days after lipofection. As for sgKvDMR#1 in **A**. **C:** Stacked histograms demonstrate the allelic mutation bias for sgKvDMR#1. Asterisks correspond to p-values for unpaired t tests on the fold-difference between maternal (red) and paternal (blue) editing frequencies for brief (16-hour) and long (4-day) exposure. **D:** Stacked histograms demonstrate the allelic mutation bias for sgInpp5f_v2. The statistics are as for sgKvDMR#1 in **C**.

4.3 Limiting Cas9 concentration results in higher allelic bias

Prolonged exposure to CRISPR/Cas9 was formerly shown to lead to similar levels mutagenesis between the two alleles in our system. To test if the concentration of Cas9 is crucial to achieve allele-specific mutagenesis I designed the following experiment. Lipofection is the preferred method to deliver the genome editing reagents in the hybrid mESCs that I am using. However, it is possible that individual cells receive different numbers of copies of the introduced plasmid, resulting in a heterogeneous population of transfected cells with variable Cas9 concentrations. In order to categorise transfected cells based on the concentration of Cas9 I used a different plasmid construct (px458, Addgene #48138, as mentioned in **Chapter 2.1.2**), where Cas9 is coupled to eGFP via a self-cleaving 2A peptide. The 2A peptide allows two proteins to be produced in a 1:1 ratio. A promoter mediates transcription of a single mRNA encoding two polypeptides, linked by the 2A sequence. As the 2A peptide emerges from the ribosome it is cleaved, liberating the first polypeptide and leaving the second polypeptide to be translated independently (De Felipe *et al.*, 2006).

Triplicate sets of both BxJ and JxB were transfected with the px458 construct including sgKvDMR#3 and the relevant ssODN. Since eGFP concentration reflects the concentration of Cas9, fluorescence activated cell sorting (FACS) was performed to sort the cells into three distinct populations according to eGFP fluorescence. The sorting was conducted at 24 hours after transfection, to allow enough expression of the eGFP fluorescent reporter, as shown for BxJ cells in **Figure 4.3.A**.

Half of the harvested cells from each population were used to isolate genomic DNA immediately after the sorting, whereas the remaining cells were maintained in culture for an additional 3 days (**Figure 4.3.B**). Therefore, data was collected for both a short and a long time of exposure to differing

concentrations of the genome editing reagents. The different populations that emerged from the sorting were maintained and analysed separately with the previously described experimental pipeline.

As shown in **Figure 4.3.C**, at low limiting Cas9 concentration the allelic bias was elevated than previously obtained values. In addition, a shorter exposure time to the genome editing reagents led to an evident increase in allelic bias, especially when the concentration of Cas9 was low. On the contrary, prolonged exposure to the genome editing reagents resulted in allelic bias values similar to the ones that were observed in previous experiments (**Figure 4.3.D**). This is true especially for the population of cells where Cas9 was abundant. In the populations with low and intermediate Cas9 levels the allelic bias was persistently higher than previously observed values, even after 4 days of exposure (**Figure 4.3.D**). One-way ANOVA followed by Tukey's HSD test were performed to assess if there are significant differences in allelic bias from individual replicates between the different cell populations with Low, Intermediate and High eGFP and Cas9 expression. Statistically significant differences are shown at the bottom panels of **Figures 4.3.B&C**.

The same experiment was also performed in parallel for the JxB hybrid line. As illustrated in **Figure 4.4.A** the experimental strategy was the same as for the BxJ hybrid line. The gating profile for JxB is shown in **Figure 4.4.B**, where cells were split and harvested once more at 24 hours and 4 days after transfection. In contrast to the BxJ line, the percentage of fully methylated reads was reduced, as assessed by illumina sequencing of bisulfite-modified DNA. This means that the mutagenesis results should be interpreted bearing in mind that there has been partial loss of imprinting for the *KvDMR* region in this hybrid mESC line. Accordingly, as shown in **Figures 4.4.C and 4.4.D**, the allelic bias is greatly reduced compared to the previously observed bias in the BxJ line.

Nonetheless, although the magnitude of effect was smaller, the same general trends were observed in the JxB line. Prolonged exposure leads to lower allelic bias values (**Figure 4.4.D**). Similar to the BxJ line, the highest

allelic bias value among the three different populations at 4 days after transfection is observed at the condition where the concentration of Cas9 is low. In contrast, for a shorter 24-hour period of exposure we observe a lack of allelic bias for low limiting Cas9 concentration, as shown on the left-hand site of **Figure 4.4.C**. This does not agree with our previous observations but could be a consequence of the aforementioned partial loss of imprinting that characterises the *KvDMR* region in the JxB line. Nevertheless, the allelic bias observed when Cas9 was abundant was lower than the equivalent value in the population with intermediate Cas9 expression, which agrees with our previous findings.

Overall, the data suggests that chromatin modifications influence the way chromatin regulates mutagenesis by CRISPR/Cas9 in a manner that relies both on the concentration of Cas9 and the time of exposure. The combination of prolonged exposure to the reagents and a high expression level of Cas9 leads to the mutagenesis reaction proceeding to completion quicker. Conversely, when Cas9 is not abundant the difference in mutagenesis between the two alleles in accentuated.

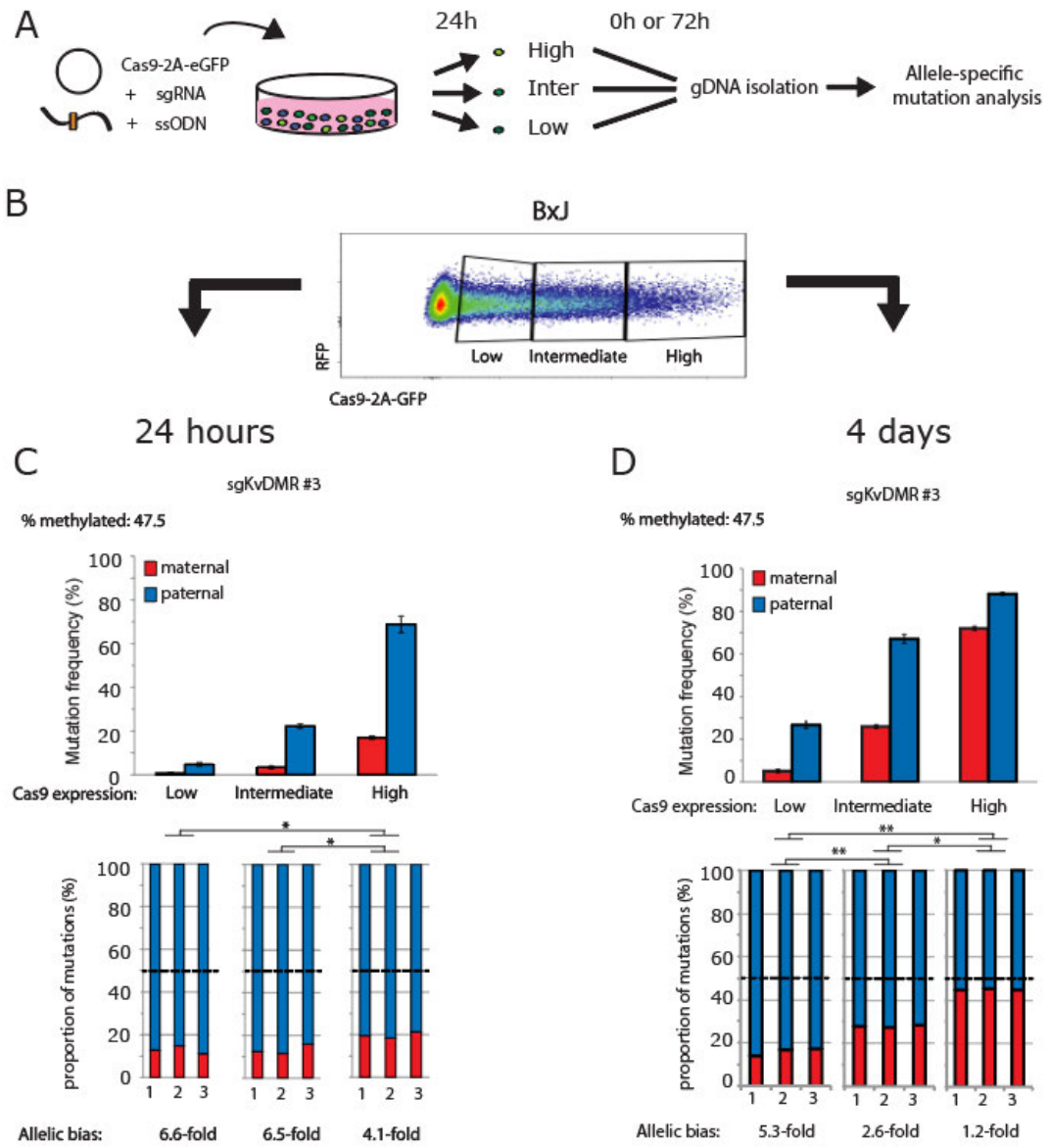


Figure 4.3 Chromatin modifications influence genome editing in a Cas9-concentration–dependent manner in BxJ hybrid mESCs. A: Schematic showing the experimental strategy for the Cas9-eGFP sorting experiment. **B:** Flow cytometry profiles demonstrating the heterogeneity in Cas9-2A-eGFP expression for the BxJ line. Data shown for sgKvDMR#3 for the promoter of *KvDMR* (position shown in **Figure 3.2.A**). **C:** Analysis at 24 hours after transfection. The percentage of methylated reads from mock-transfected cells without selection for Cas9 expression is shown at the top. Imprinting was highly maintained in this line. Error bars represent the Standard Deviation of three experimental replicates. Stacked histograms at the bottom illustrate the allele-specific mutation bias in each population. Asterisks for Tukey's HSD test p-values per pairwise comparisons *$p < 0.05$, **$p < 0.01$. **D:** Analysis at 4 days after transfection, as described above for 24 hours after transfection.

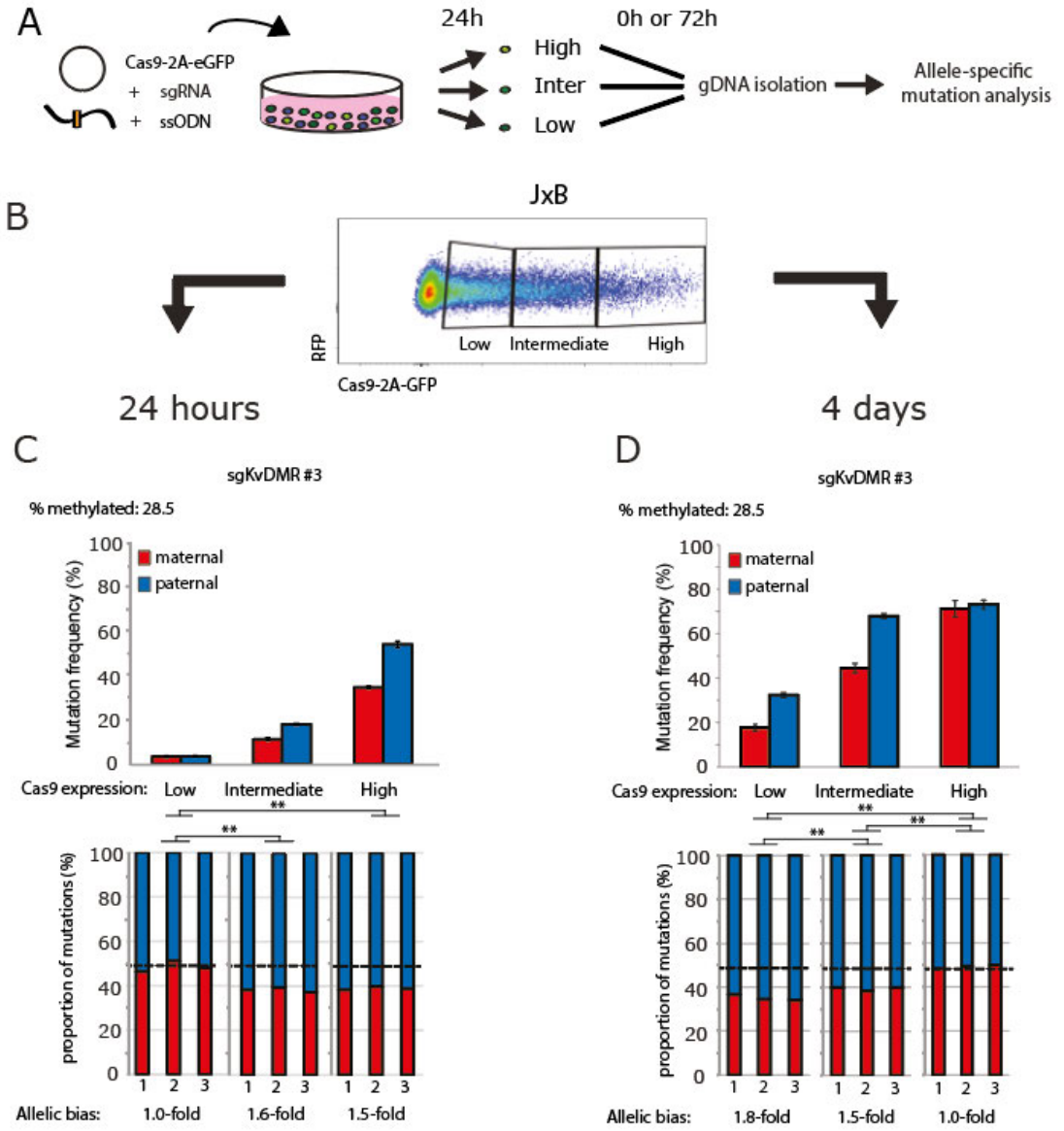


Figure 4.4 Chromatin modifications influence genome editing in a Cas9-concentration–dependent manner in JxB hybrid mESCs. A: Schematic showing the experimental strategy for the Cas9-eGFP sorting experiment. **B:** Flow cytometry profiles demonstrating the heterogeneity in Cas9-2A-eGFP expression for the JxB line, which was similar to the BxJ line. Data shown for sgKvDMR#3 for the promoter of *KvDMR* (position shown in **Figure 3.2.A**). **C:** Analysis at 24 hours after transfection. The percentage of methylated reads was lower compared to BxJ, indicating partial loss of imprinting. Error bars represent the Standard Deviation of three experimental replicates. Stacked histograms at the bottom illustrate the allele-specific mutation bias per population. Asterisks for Tukey's HSD test p-values per pairwise comparisons *$p < 0.05$, **$p < 0.01$. **D:** Analysis at 4 days after transfection, as described above for 24 hours after transfection.

4.4 Accessible chromatin allows higher levels of dCas9 occupancy

Having identified two factors that regulate the effect repressive chromatin modifications exert on targeted mutagenesis with CRISPR/Cas9, we then decided to gain further insights into the inhibition mechanism. Previous reports have shown that catalytically dead Cas9 (dCas9) is able to associate with many off-target sites, the majority of which is found in open chromatin regions. These findings have emerged from genome-wide chromatin immunoprecipitation and sequencing (ChIP-Seq) studies in mESCs (Wu *et al*., 2014) and HEK293 cells (Kuscu *et al.*, 2014). In order to ascertain at which stage of the CRISPR/Cas9 mutagenesis procedure chromatin exerts an effect we adapted a similar approach to assess the binding specificity of dCas9 in our previously established system in the hybrid mESCs in a time-dependent manner.

In general, SpCas9 must first associate with the sgRNA molecule in order to begin searching for its target site in the genome (Jiang *et al.*, 2015). Upon recognition of an NGG protospacer-adjacent motif (PAM) nucleotide triplet, SpCas9 stops and interrogates the adjacent sequence. Then, if the latter is complementary to the protospacer on the sgRNA molecule then it is identified as the appropriate target site (Sternberg *et al.*, 2014). Evidence from a recent study suggests that the complex diffuses in a lateral manner on the DNA while searching for PAM triplets (Globyte *et al.*, 2019). Nevertheless, all the above conclusions have been drawn through the use of *in vitro* systems and fail to consider the effect of chromatin modifications on the association rate of the assembled Cas9-sgRNA complex on mammalian genomes *in vivo*.

To investigate this, I performed a time course experiment to assess the degree of binding of catalytically dead Cas9 (dCas9) in an allele-specific manner. After harvesting at different timepoints following transfection, cells were crosslinked with formaldehyde and ChIP was performed followed by

qPCR and targeted amplicon NGS. As illustrated in the diagram in **Figure 4.5.A** cells were transfected with plasmid constructs encoding the relevant sgRNAs and dCas9 with a triple flag tag (3xFLAG-dCas9). After crosslinking and sonication was performed, sheared chromatin was incubated with an anti-flag antibody to enrich for genomic regions bound to dCas9.

Enrichment analysis by qPCR revealed that the highest level of dCas9 association on the *KvDMR* and promoter region of *Impact* was at 16 hours following lipofection, as shown in **Figure 4.5.B** and **Figure 4.5.C**, respectively. Target amplicon sequencing and allele-specific read analysis revealed that dCas9 shows approximately two-fold greater occupancy on the accessible paternal allele than the imprinted maternal allele for both regions.

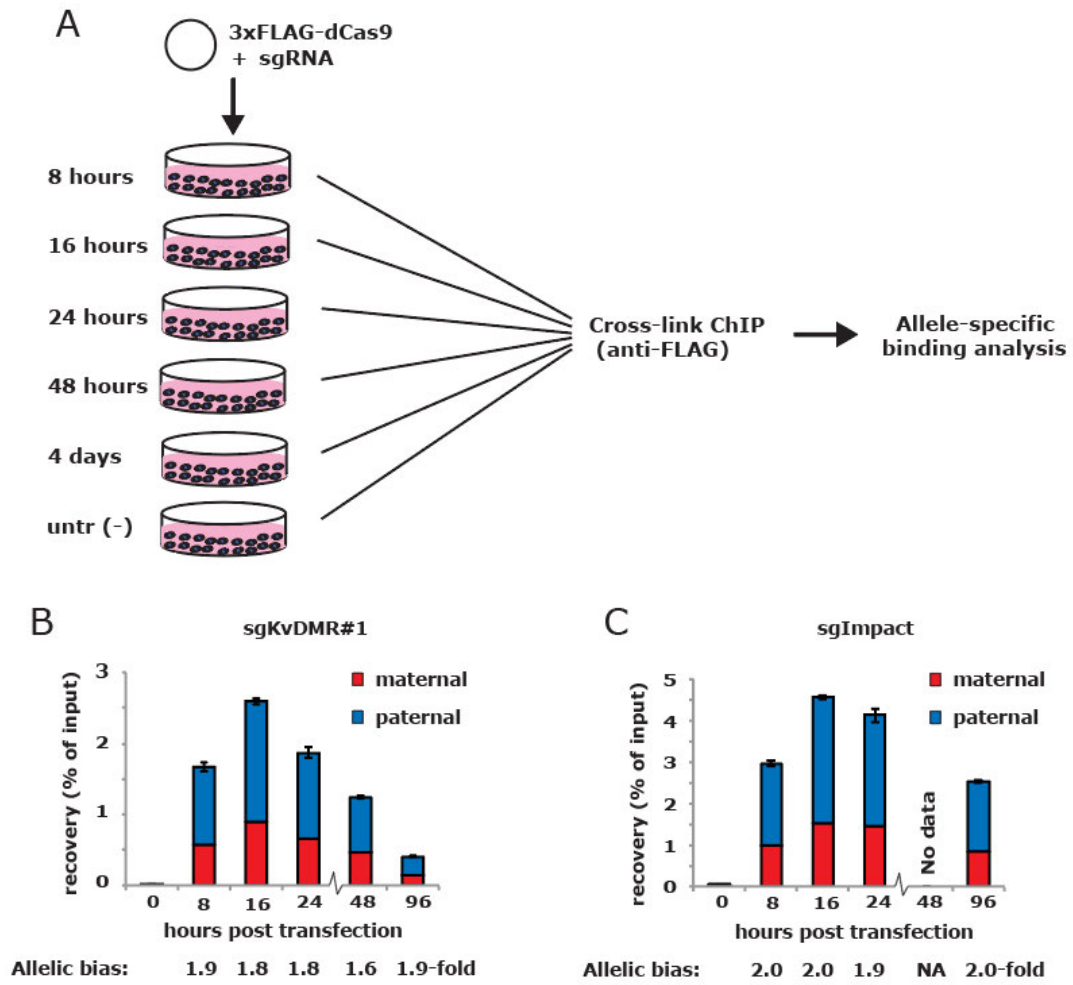


Figure 4.5 Enhanced Cas9 occupancy in accessible compared to inaccessible chromatin. A: Schematic showing the experimental strategy for the dCas9 time course ChIP experiment. This was only performed once and only in the BxJ line. **B:** Stacked histograms showing allele-specific enrichment of dCas9 on the *KvDMR* region, expressed relative to Input DNA. Error bars represent the Standard Deviation of three technical qPCR replicates. Allelic bias was calculated from subsequent targeted PCR amplicon sequencing. **C:** Stacked histograms showing allele-specific enrichment of dCas9 on the promoter region of *Impact*, expressed relative to Input DNA. Error bars represent the Standard Deviation of three technical qPCR replicates. Allelic bias was calculated from targeted PCR amplicon sequencing.

4.5 Studying the effect of repressive chromatin modifications during off-target mutagenesis by CRISPR/Cas9

In general, off-target mutagenesis is a major problem during genome editing with CRISPR/Cas9 but may be subject to different constraints relative to on-target due to the presence of mismatches between guide and target. Emerging evidence from another *in vitro* study suggests that off-target mutagenesis by CRISPR/Cas9 is inhibited by nucleosome positioning, even if the target site is at nucleosome exit site (Hinz, Laughery and Wyrick, 2016). It was also suggested that the position of the mismatch is critical for the degree of inhibition. This study showed that if the mismatch located within positions 1-10 out of 20 in the protospacer, which is proximal to the PAM site, then the effect was reported to be greater. Similar conclusions about the position of the mismatch have been drawn during a study in HEK293 cells (Hsu *et al.*, 2013), which did not assess how chromatin may further affect mismatch tolerance during off-target site recognition.

To experimentally measure this effect using the internally controlled system that we established, we modified the sgRNAs that were previously used on maternally-imprinted loci to contain mismatches on certain positions within the protospacer sequence. This would perturb full complementarity between the protospacer and the target site for two of the maternally imprinted regions we have previously characterised and targeted with perfectly-matching sgRNAs. I focused on the *KvDMR* region and the promoter region of *Impact*, due to the frequent loss of imprinting that I previously observed on the promoter region of *Inpp5fv2* (**Figure 3.9** and **Figure 3.10.B**). Finally, I also decided to compare the effect of different combinations of mismatches within the protospacer sequence of the sgRNA.

All the experiments with mismatch-containing sgRNAs were performed in a 4-day timeframe with puromycin selection, similar to the very first editing

experiments shown in **Figure 3.7** and **Figure 3.8**. Since the corresponding perfect-match sgRNAs were not used to edit separate cell populations in parallel with the mismatch-containing sgRNA editing experiments, the data shown in **Figures 4.7-4.12** is missing an optimal control to compare the effect exerted by repressive chromatin modifications for on-target versus off-target mutagenesis. Therefore, throughout sections **4.5.1**, **4.5.2** and **4.5.3** the data from **Figure 3.7.A** and **Figure 3.8.A** is used as a suboptimal control for on-target mutagenesis for the perfect match sgKvDMR#1 and sgImpact, respectively.

4.5.1 Single mismatches and positioning within the protospacer

Firstly, I assessed the effect of single nucleotide mismatches within the protospacer. Hinz, Laughery and Wyrick (2016) reported that the closer the mismatch is located to the PAM site, the more likely it is to affect the cleavage efficiency of Cas9. This effect was specific to mismatches on positions 1 to 10 on the protospacer. Therefore, I compared the effect of a single mismatch when it is found either within or outside of this region of the protospacer. In addition, a single mismatch at position 14 led to an overall decrease in the cleavage potency of the sgRNA. Therefore, I used sgRNAs that bear a single nucleotide substitution on position 1 or position 14. The results for sgKvDMR#1 and sgImpact are shown in **Figure 4.7** and **Figure 4.8**, respectively.

As for previous targeting experiments, I transfected the BxJ and JxB hybrid mESC lines with mismatch-bearing sgRNAs in the px459v2 vector. The time of exposure for all experiments with sgRNAs that contain mismatches was 4 days, to include puromycin selection of successfully transfected cells. When the first position of the sgKvDMR#1 protospacer was substituted (**Figure 4.7.A**) the overall potency of sgKvDMR#1 dropped below 10% (**Figure 4.7.B**),

compared to 80% for sgKvDMR match (**Figure 3.7.A**). In contrast, when the same position was mutated in the protospacer of sgImpact (**Figure 4.8.A**) the overall editing efficiency of sgImpact mm1, shown in **Figure 4.8.B**, was similar to that of the perfect sgImpact match (**Figure 3.8.A**). The same was true for both sgKvDMR#1 and sgImpact when the nucleotide on position 14 of the protospacer was mutated (**Figure 4.7.C** and **Figure 4.8.C**).

An increase in allelic bias, compared to the perfect match sgRNAs, was observed for sgKvDMR#1 mm1 and sgImpact mm1 in BxJ, where imprinting was better maintained and there was an overall reduction in editing by a mismatch-containing sgRNA, relative to the perfect match sgRNA (**Figure 4.7.B**). In this case, the paternal non-imprinted allele is edited more frequently than the maternal allele. In contrast, sgKvDMR#1 mm14 (**Figure 4.7.C**) and sgImpact mm14 (**Figure 4.8.C**) did not yield any increase in allelic bias compared to the equivalent perfectly matching sgRNAs for both target loci. This raises the possibility that when the overall editing potency of the sgRNA decreases allelic bias increases, implying that there is a chromatin-related effect.

It is possible that chromatin modifications do not influence off-target site recognition when there are only single mismatches within the protospacer. A profound effect is only observed for sgKvDMR#1 mm1, where the first position next to the PAM site is mutated and the overall editing efficiency of the mismatch containing sgRNA is reduced. However, as shown in **Figure 4.8.B** the effect is more subtle for sgImpact mm1. This could be either owed to sequence-specific properties of the two different sgRNAs or due to the different degrees of imprinting maintenance that were observed from the percentage of methylated reads for each target locus at the time the experiment was performed. Additionally, a distal single nucleotide mutation at position 14 was not sufficient to increase the allelic bias in either of the two target loci. Therefore, it is possible that the overall mutation efficiency and position of the single mismatch within the protospacer affect the inhibition of off-target site recognition on the maternal allele.

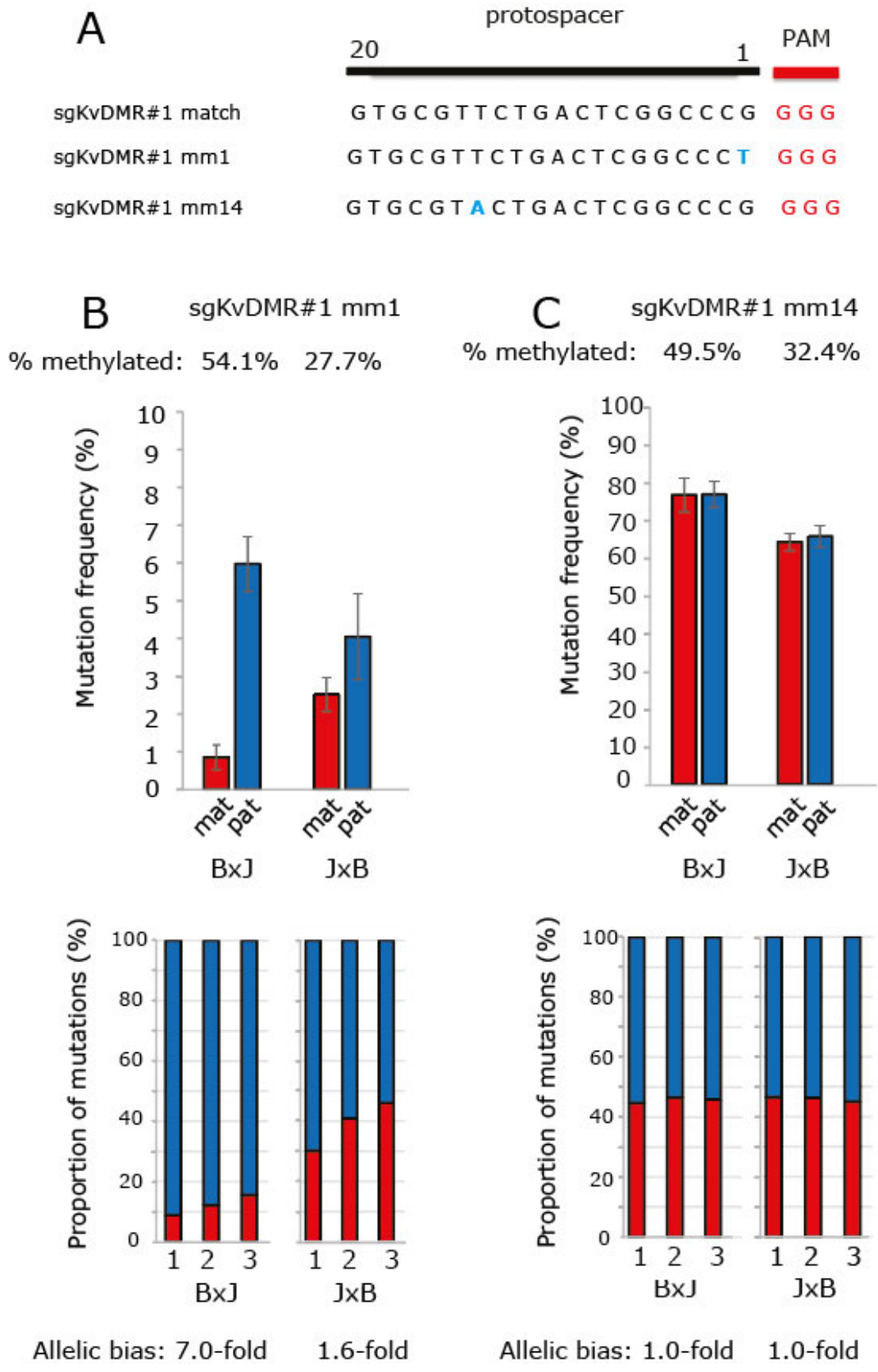


Figure 4.6 A PAM-proximal single mismatch influences target site recognition for sgKvDMR#1.

Figure 4.6 A PAM-proximal single mismatch influences target site recognition for sgKvDMR#1. A: Schematic of the sequences for the perfect match sgKvDMR#1 and the two single mismatch-containing sgRNAs. The single nucleotide substitutions are shown in light blue in relation to the PAM site, shown in red. mm = mismatch **B:** Allele-specific editing frequencies for sgKvDMR#1 mm1. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above every histogram to show the degree of imprinting maintenance during each targeting experiment. Stacked histograms that demonstrate the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgKvDMR#1 mm14. As usual, error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

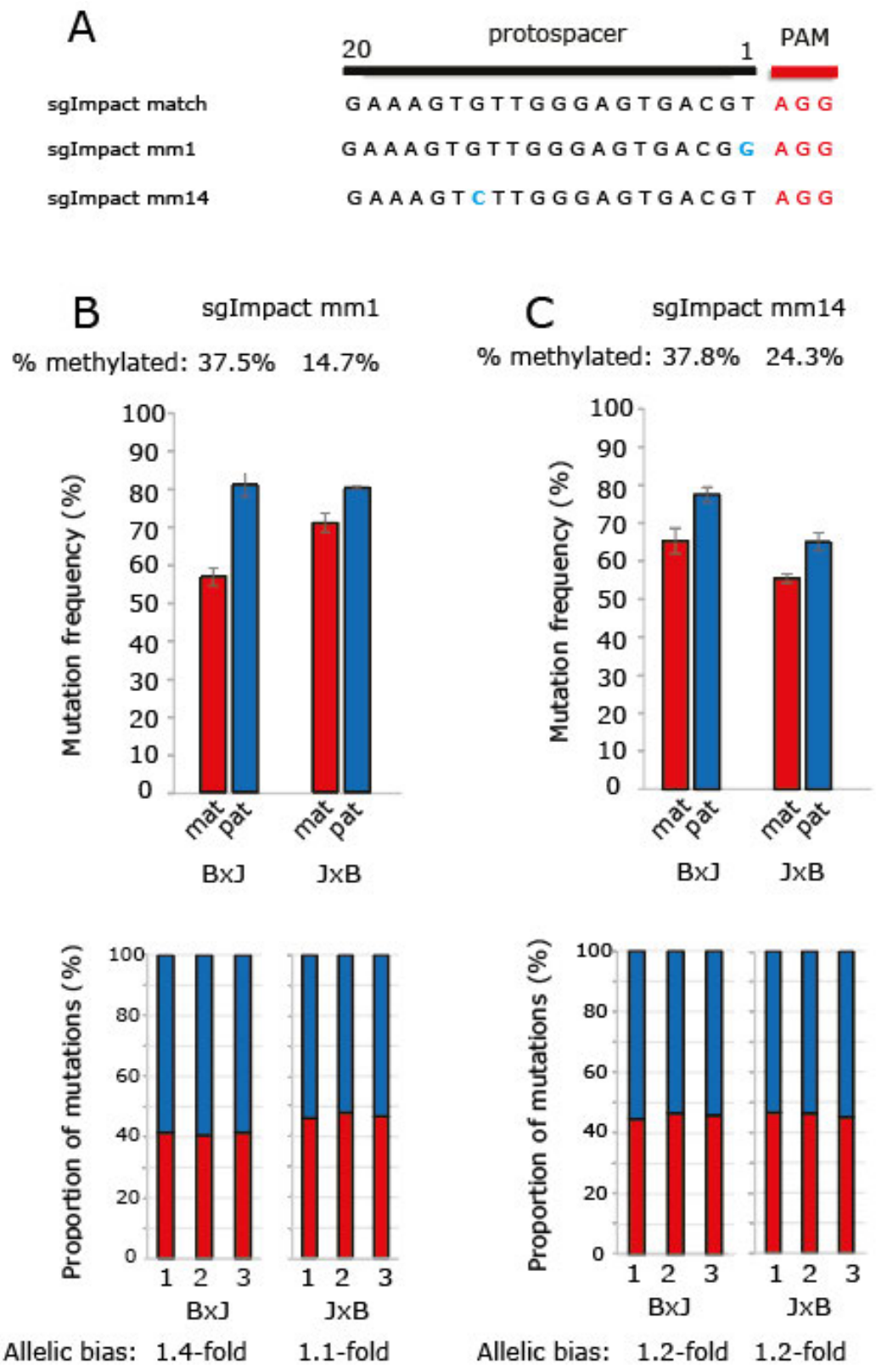


Figure 4.7 Single nucleotide mismatches do not influence target site recognition for sgImpact.

Figure 4.7 Single nucleotide mismatches do not influence target site recognition for sgImpact. A: Schematic of the sequences for the perfect match sgImpact and the two single mismatch-containing sgRNAs. The single nucleotide substitutions are shown in light blue in relation to the PAM site, shown in red. mm = mismatch **B:** Allele-specific editing frequencies for sgImpact mm1. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides shown at the top of the histogram shows the degree of imprinting maintenance during this targeting experiment. Stacked histograms that demonstrate the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgImpact mm14. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

4.5.2 Two mismatches and positioning within the protospacer

The single mismatch experiments from the previous section suggested that chromatin disproportionately affects mutagenesis in the presence of mismatches, but the effect is only observed when there is an overall reduction in mutagenesis. To explore this further, we introduced two mismatches to increase the likelihood that the overall mutation frequency would decrease. As for single mismatches, I compared the effect of two PAM-proximal or PAM-distal mismatches. This time I introduced mismatches in positions 1 and 10, which are proximal to the PAM site, and positions 11 and 12, which are distal. The nucleotide substitutions are illustrated in **Figure 4.9.A** for sgKvDMR#1 and in **Figure 4.10.A** for sgImpact.

A reduction in the overall editing potency of both sgKvDMR#1 and sgImpact is evident after the introduction of two mismatches in the protospacer. The reduction is disproportionate and related to the maternal inaccessible allele. The effect is more profound for sgKvDMR#1 mm1+10, where the highest allelic bias is observed (**Figure 4.9.B**). Distal mismatches on positions 11 and 12 also raised the allelic bias (**Figure 4.9.C**), though not as much as the combination of two proximal mismatches in positions 1 and 10.

A similar trend is also present for sgImpact. This is only observed in BxJ, since a great reduction in DNA methylation was observed at the promoter region of *Impact* in JxB during this set of experiments. The combination of mismatches 1 and 10 increased the bias between off-target site recognition towards the accessible paternal allele. It is also striking that this double mismatch combination also reduced the overall potency of sgImpact to levels below 10% for both alleles (**Figure 4.10.B**). As shown in **Figure 4.10.C**, the overall potency of sgImpact mm11+12 was not reduced as drastically as of sgImpact mm1+10 and the observed rise in allelic bias of sgImpact mm11+12 was not as profound as for sgImpact mm1+10.

Overall, the presence of two mismatches within the protospacer results in elevated inhibition of mutagenesis on the maternal allele compared to a single mismatch. The position of the mismatches is also critical, since the closer they are located to the PAM site, the more the effect is accentuated.

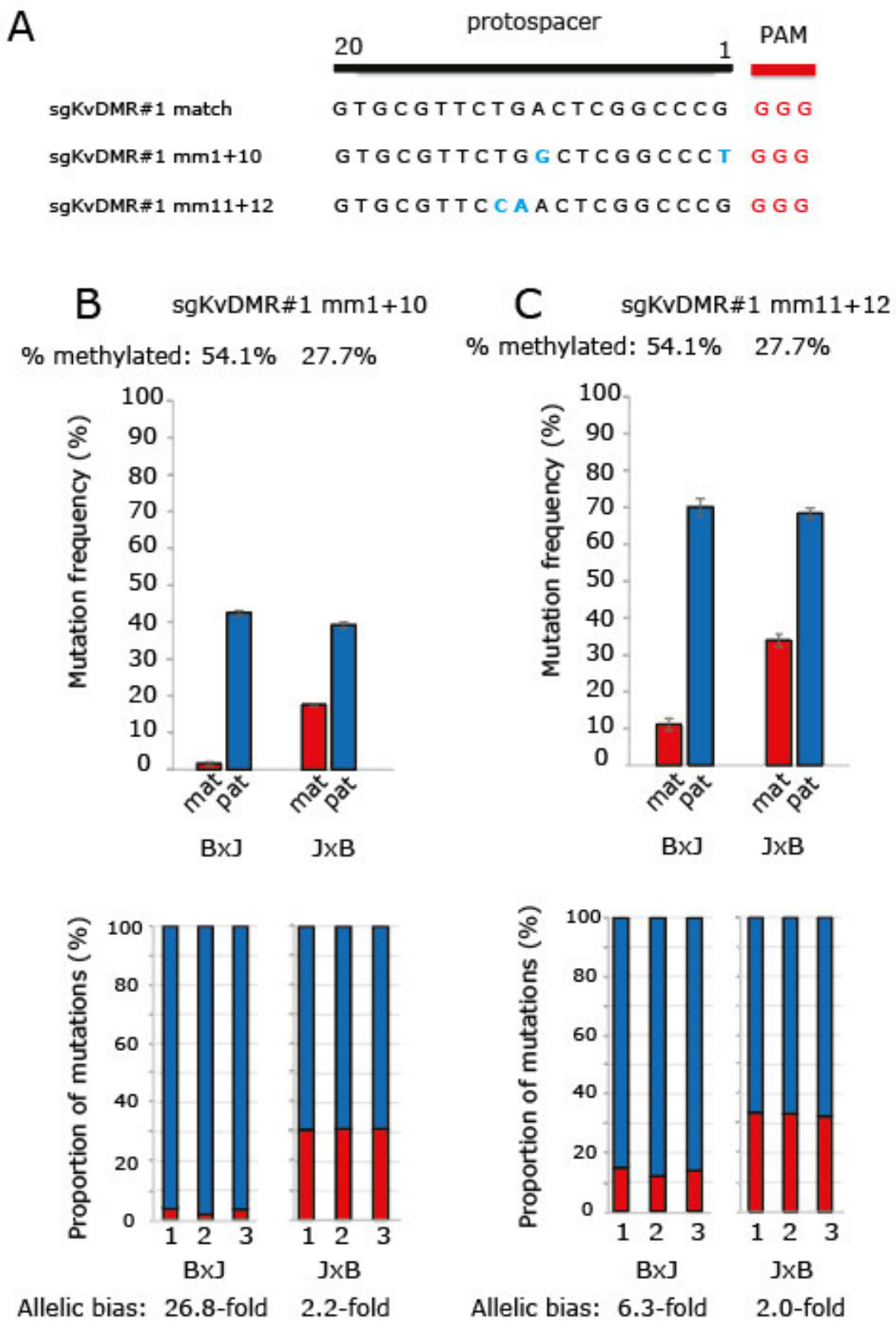


Figure 4.8 Two mismatches in sgKvDMR#1 disproportionately reduce target site recognition on the paternal allele.

Figure 4.8 Two mismatches in sgKvDMR#1 disproportionately reduce target site recognition on the paternal allele. A: Schematic of the sequences for the perfect match sgKvDMR#1 and the two combinations of double mismatches in sgKvDMR#1. The individual nucleotide substitutions for each position are shown in light blue in relation to the PAM site, shown in red. **B:** Allele-specific editing frequencies for sgKvDMR#1 mm1+10. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above every histogram to show the degree of imprinting maintenance during this experiment. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgKvDMR#1 mm11+12. As for above, error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

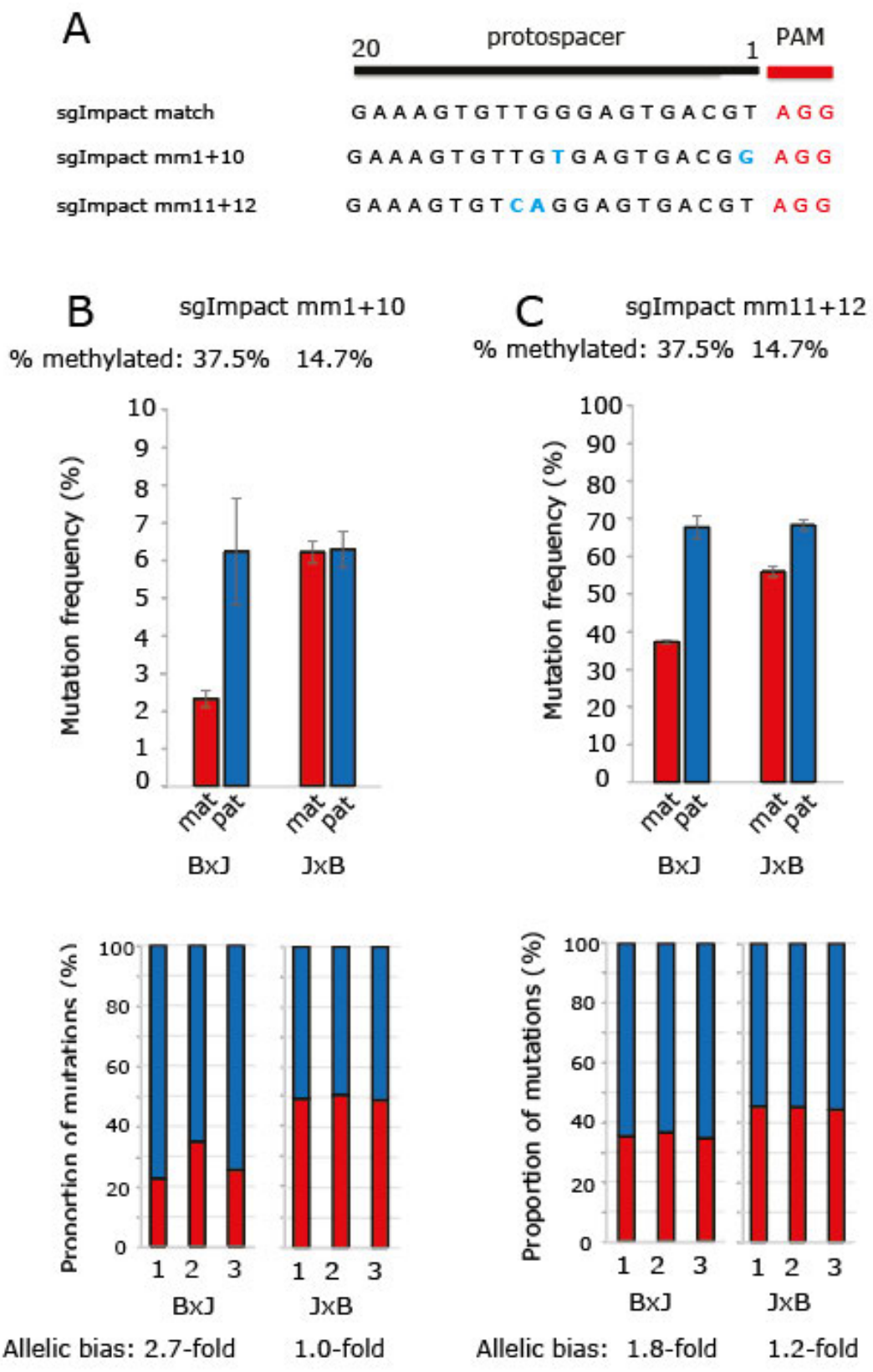


Figure 4.9 Subtle reduction of target site recognition on the maternal allele when sgImpact bears two mismatches.

Figure 4.9 Subtle reduction of target site recognition on the maternal allele when sgImpact bears two mismatches. Schematic of the sequences for the perfect match sgImpact and the two combinations of double mismatches in sgImpact. The individual nucleotide substitutions are shown in light blue in relation to the PAM site, shown in red. **B:** Allele-specific editing frequencies for sgImpact mm1+10. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides shown at the top of the histogram shows the degree of imprinting maintenance during this targeting experiment. Stacked histograms that demonstrate the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgImpact mm11+12. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

4.5.3 Presence of a third mismatch within the protospacer

In the end, I also investigated the effect of chromatin modifications on off-target site recognition when more than two nucleotide mismatches are present within the protospacer. In order to avoid reducing the overall editing potency of the sgRNAs to a drastic extend by introducing three mismatches near the PAM site, positions 14, 11 and 8 were selected after examination of a single-nucleotide specificity matrix that was published for Cas9 (Hsu *et al.*, 2013). Therefore, two out of three of the selected mismatches were introduced on the second half of the protospacer. I also assessed the combined effect of these two distal mismatches for positions 14 and 11 in comparison to the triple mismatch-containing sgRNAs for both sgKvDMR#1 and sgImpact.

Interestingly, these mismatch combinations led to extreme reductions to the potency of the sgRNAs as shown in the histograms throughout **Figure 4.11** and **Figure 4.12**. Notably the overall mutation frequency on the imprinted target sites was around 1% or lower, which was the lowest of all the different combinations of mismatches that were tested previously.

Despite the greatly reduced overall mutation frequency resulting from these nucleotide mismatches, a reduction was observed on the maternal imprinted allele. More specifically, **Figure 4.11.B** demonstrates that the majority of sgKvDMR#1 mm14+11 edits occurred on the paternal allele. This trend is also true for sgKvDMR#1 mm14+11+8, thought the effect is more subtle due to the presence of the third mismatch at position 8 (**Figure 4.11.C**).

The results for sgRNA targeting the promoter region of *Impact* are once more similar. **Figure 4.12.B** shows that sgImpact mm14+11 mostly generated mutations on the paternal allele, though it is evident from the allelic bias value that the effect is more subtle compared to sgKvDMR#1 mm14+11 from **Figure 4.11.B**. In contrast, introducing a third mismatch on position 8 of sgImpact did not reduce the mutation frequency on the maternal allele for this target locus (**Figure 4.12.C**).

Overall, the low mutation frequencies observed for these combinations of mismatches should be interpreted with caution. It is also not clear if the differences we observed between the two loci are owed to sequence-specific properties of the two different sgRNAs or the presence of a third mismatch within the protospacer.

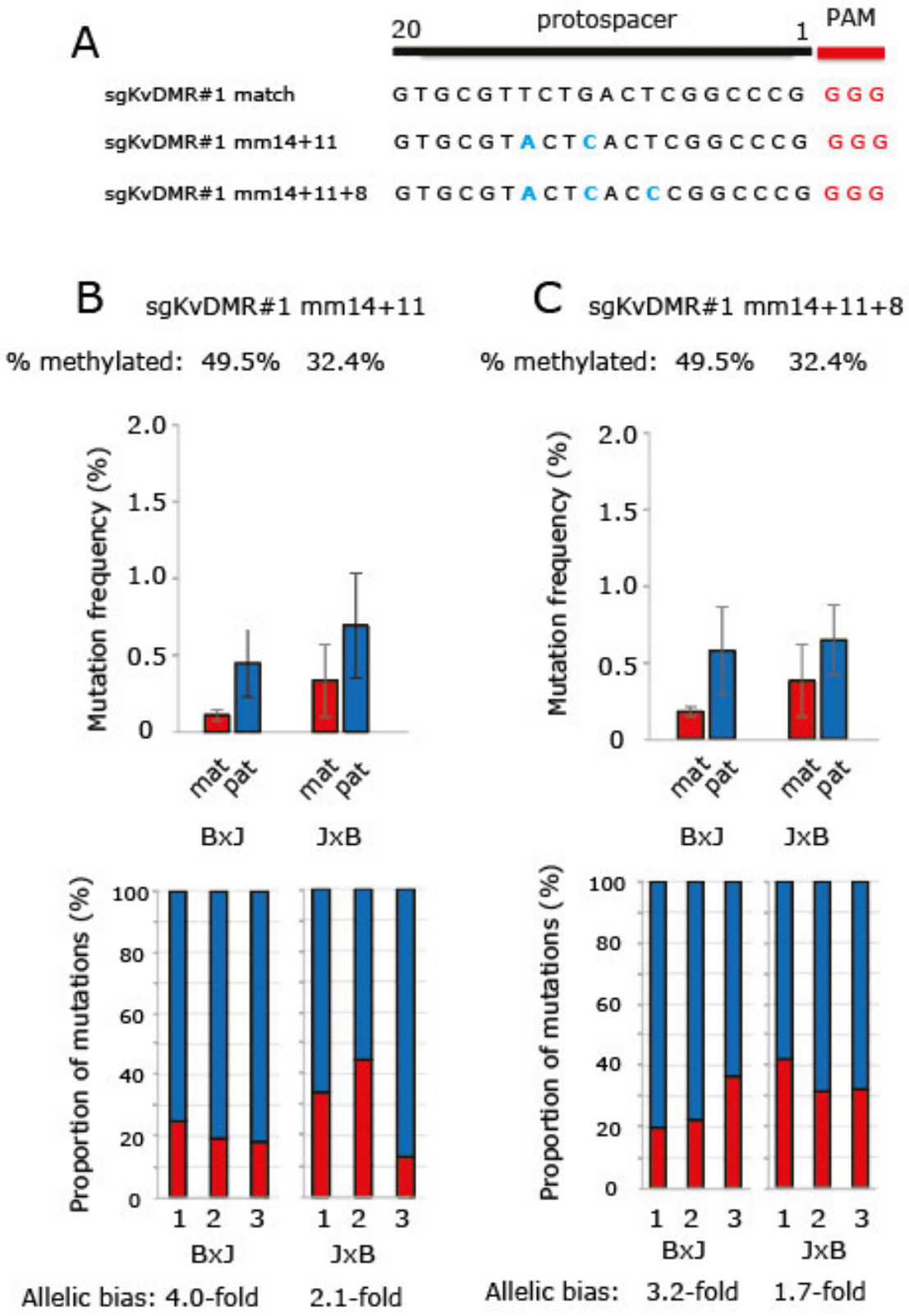


Figure 4.10 A third mismatch in sgKvDMR#1 does not contribute to the allele-specific inhibition of off-target mutagenesis.

Figure 4.10 A third mismatch in sgKvDMR#1 does not contribute to the allele-specific inhibition of off-target mutagenesis. A: Schematic of the sequences for the perfect match sgKvDMR#1 and the mismatches in positions 14, 11 and 8. The individual nucleotide substitutions for each position are shown in light blue in relation to the PAM site. **B:** Allele-specific editing frequencies for sgKvDMR#1 mm14+11. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram to illustrate the degree of imprinting maintenance during this experiment. Stacked histograms that demonstrate the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgKvDMR#1 mm14+11+8. As for above, error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

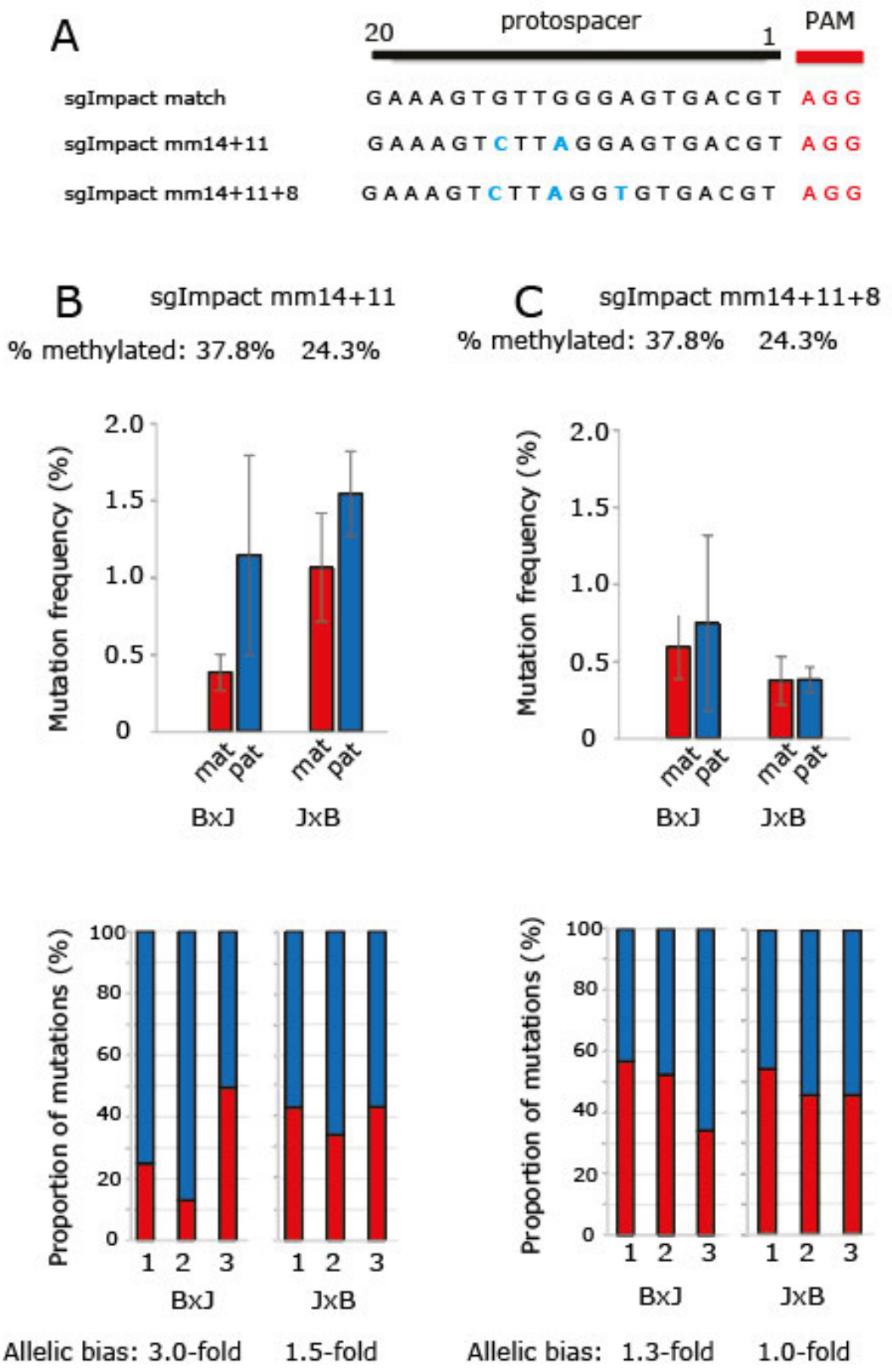


Figure 4.11 A third mismatch in sgImpact does not affect allele-specific inhibition of off-target mutagenesis.

Figure 4.11 A third mismatch in sgImpact does not affect allele-specific inhibition of off-target mutagenesis. Schematic of the sequences for the perfect match sgImpact and the mismatches in positions 14, 11 and 8. The individual nucleotide substitutions are shown in light blue in relation to the PAM site. **B:** Allele-specific editing frequencies for sgImpact mm14+11. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides shown at the top of the histogram shows the degree of imprinting maintenance during this targeting experiment. Stacked histograms that demonstrate the allelic mutation bias are shown at the bottom. **C:** Allele-specific editing frequencies for sgImpact mm14+11+8. Error bars represent the Standard Deviation of three experimental replicates. The percentage of reads with fully methylated CG dinucleotides is shown above the histogram. Stacked histograms demonstrating the allelic mutation bias are shown at the bottom.

4.6 Discussion

In this chapter I described how I identified additional factors that strongly influence how chromatin and epigenetic modifications regulate CRISPR/Cas9 mutagenesis. Assessing allele-specific mutagenesis for different exposure times revealed that mutations appear more rapidly on the paternal euchromatic allele.

Prolonged exposure to the genome editing reagents allowed mutations on the maternal heterochromatin-associated allele to accumulate over time, leading to almost equal mutation frequencies after 4 days of exposure. The concentration of Cas9 was also important towards allele-specific genome editing. High levels of Cas9 expression led to a faster completion of the mutagenesis process during 4 days of targeting. In contrast, the lower the concentration of Cas9, the higher the editing bias for the paternal euchromatic allele. These findings are novel and will contribute towards improving our current understanding about how CRISPR/Cas9 is affected by eukaryotic chromatin compaction.

The prospect that CRISPR/Cas9 and related genome editing tools will be used to treat human disease in the near future looks increasingly likely, but should be regarded with care and caution (Baltimore *et al.*, 2015). Recent successful examples of the application in human hematopoietic stem/progenitor cells (hHSPCs) highlight the potential of using this tool in a clinical context (DeWitt *et al.*, 2016; Chung *et al.*, 2019). Additionally, the use of improved engineered versions Cas9 on synchronised human hematopoietic stem cells (hHSCs) boosts the genome editing efficiency in this primary cell line through temporal control of DNA repair (Lomova *et al.*, 2019).

In general, it is desirable to have a high on-target editing efficiency with as few off-target mutations as possible. Our results suggest a strategy to use a priori knowledge of chromatin state to achieve this. When the intended target site is within accessible chromatin, but potential off target sites are not, brief

exposure to Cas9 in conjunction with maintaining the expression level of Cas9 low would favour on versus off-target mutagenesis to a greater extent than result in the accumulation of edits on the appropriate target site. In addition, the chromatin context of predicted off-target sites should also be considered before choosing to use a sgRNA. All the experiments performed with mismatch-containing sgRNAs (**Chapter 4.5**) were 4-days long and included puromycin treatment selection of successfully-transfected cells. Despite the prolonged exposure to CRISPR/Cas9, we demonstrated that off-target mutagenesis is reduced within inaccessible heterochromatic regions when the overall editing efficiency is also reduced.

In the past, DNase hypersensitive regions were deemed more permissive for dCas9 to associate with an off-target site (Kuscu *et al*., 2014). A more recent study has proved this is also true for wild-type Cas9 after assessing several different combinations of single or double mismatches within the protospacer (Kim and Kim, 2018). In contrast to the above, our approach does not solely rely on accessibility based on DNase hypersensitivity but also on additional repressive epigenetic marks associated with constitutive heterochromatin, such as DNA methylation and H3K9me3.

While performing the experiments that are described in **Chapter 4.5**, a study was published where it was argued that inaccessible chromatin regions are less prone to off-target CRISPR/Cas9 mutagenesis (Kim and Kim, 2018). This time two sgRNAs were assessed in HEK293 and HeLa cells, in a system where mismatch-containing sgRNAs were used to target two sequences that were present in two copies in the human genome. However, the two target sites were selected based on DNase hypersensitivity alone in the absence of other epigenetic modification differences between the two copies of the target sequences, such as DNA methylation and histone modifications. This encouraged me to continue investigating the effect of repressive chromatin modifications on off-target mutagenesis in our system.

We also demonstrated that different combinations of mismatches within the protospacer lead to different degrees of inhibition of off-target mutagenesis

within heterochromatic regions, as measured by the allelic bias ratio. Kim and Kim (2018) showed that several combinations of double mismatches in the protospacer lead to the biggest differences in target site recognition between accessible and inaccessible regions, whereas single nucleotide mismatches led to more subtle differences. This agrees with our observation that some combinations of double mismatches generate higher allelic bias than single mismatches. However, great variability was also observed when more than one target sites were interrogated, since every detectable effect was more subtle for sgImpact compared to sgKvDMR#1. The sequence and secondary structure of different sgRNAs determines their overall efficiency (Moreno-Mateos *et al*., 2015; Jensen *et al*., 2017). Therefore, it is likely that the sgRNAs we used possess different properties that are determined by the different protospacer sequences. This highlights the need for a more systematic characterisation and assessment of more mismatch combinations for multiple target sites.

An earlier study on the specificity of CRISPR/Cas9 showed that while the sequence context and the position of mismatches is important for on-target site recognition, specific DNA modifications like DNA methylation did not affect the editing activity of CRISPR/Cas9 using an *in vitro* system (Hsu *et al.*, 2013). A different study using an *in vivo* system later showed that DNA methylation did not affect the initial association of CRISPR/Cas9 with its target site (Fujita, Yuno and Fujii, 2016). The data presented in **Figure 4.5** in this chapter indicate that the presence of repressive epigenetic modifications hinder the initial association of CRISPR/dCas9 with its target site. However, these data are limited to two sgRNAs. Given the diverse effect that DNA methylation has on the binding of transcription factors to their destined DNA sequence from mammals to plants (O'Malley *et al.*, 2016; Yin *et al.*, 2017; Héberlé and Bardet, 2019), more systematic assessment of the effect of DNA methylation on CRISPR/Cas9 binding should be performed through the use of simple *in vitro* systems in the future.

Genome editing is more permissive in euchromatic regions due to the higher level of Cas9 occupancy. In 2013 Wu *et al* reached a similar conclusion

after they performed genome-wide dCas9 ChIP-Seq analysis. Dissecting the mechanism through which repressed chromatin disproportionately impairs mutagenesis for sgRNAs containing mismatches would help us identify the causative chromatin modification and contribute towards building better tools for off-target site prediction in the future.

Chapter 5

Assessing the effect of pre-existing chromatin modifications during repair of Cas9-induced DSBs

5.1 Introduction

Having shown that chromatin compaction affects the efficiency of CRISPR/Cas9 mutagenesis we were interested to investigate how pre-existing repressive chromatin modifications of a genomic locus affect the outcome of genome editing by CRISPR/Cas9.

Double-strand breaks (DSBs) by the Cas9 endonuclease are usually repaired through two very different DNA repair mechanisms (Doudna and Charpentier, 2014). The first is non-homologous end joining (NHEJ), which is favoured throughout the cell cycle (Mladenov *et al.*, 2016) and can create a mutational scar at the CRISPR/Cas9 target site due to the generation of InDels. The second mechanism is homology-directed repair (HDR), which resembles SSA and relies on co-delivery of externally provided repair templates (Doudna and Charpentier, 2014). Despite its temporal restriction through the cell cycle and lower frequency of occurrence compared to NHEJ within mammalian cells, HDR is the coveted DNA repair pathway during genome editing with CRISPR/Cas9.

The chromatin environment and spatial positioning in the nucleus has been reported to affect DNA repair pathway choice of DSBs that are not generated by the Cas9 endonuclease in mammalian cells (Lemaître *et al.*, 2014; Burman *et al.*, 2015; Clouaire and Legube, 2015). Most notably, epigenetic modifications, like the H3K36me3 mark, have been reported to play a role towards the recruitment of factors involved in HR (Aymard *et al.*, 2014). However, very little is known about how pre-existing epigenetic chromatin modifications influence HDR frequencies and the InDel spectrum across different target sites during genome editing. We therefore decided to use our genomic imprinting-based system to gain useful insights in this previously uncharacterised area.

The experimental work presented in this chapter was primarily my own. Alison Meynert and Tracy Ballinger conducted the bioinformatic analysis, as

described in Materials and Methods, and additionally recorded the different types of InDels for the allele-specific profiling.

5.2 Repressive chromatin modifications do not affect the efficiency of homology-directed repair during genome editing

In order to report on the outcome of mutagenic DNA repair in the context of genome editing with CRISPR/Cas9, we assessed the mutation outcome for the five imprinted target sites we studied so far. An exogenous single-stranded oligodeoxynucleotide (ssODN) template was used as a donor, allowing us to distinguish HDR events. The ssODN for each target site contains a specific sequence signature that differs from the original target site sequence such that the PAM motif is abolished. Upon introduction of the information included in the donor template through HDR, subsequent rounds of Cas9 cleavage are therefore prevented. This allows us to derive the frequency of HDR events, relative to mutagenic NHEJ, in an allele-specific manner.

As described in the **Figure 3.5**, the parental origin of all reads is determined first by the presence of informative SNPs within the BxJ and JxB hybrid mESC lines. Alison Meynert then used a custom script to identify HDR reads, which were the reads that contained the sequence change we designed to abolish the PAM site from the externally provided ssODN. HDR reads were divided over the total number of reads that contained evidence of CRISPR/Cas9 mutagenesis, which were reads with evidence of NHEJ (NHEJ reads) and HDR reads. This process was performed separately for every replicate of each sample. The average of three experimental replicates was

calculated and used to report the HDR frequency for each sgRNA in either of the two hybrid mESC lines that were used.

All results illustrated in **Figure 5.1** correspond to edited genomic DNA samples collected 4 days post transfection. In general, HDR frequencies are very similar between the closed maternal allele and the accessible paternal allele for all sgRNAs across the *KvDMR* region (**Figure 5.1.A**) and the sgRNAs for the promoter regions of *Inpp5f_v2* (**Figure 5.1.B**) and *Impact* (**Figure 5.1.C**). A subtle significant preference for HDR occurrence on the euchromatic paternal allele was observed for sgKvDMR#2 in the BxJ hybrid mESC line (**Figure 5.1.A**, centre), but the absence of consistency in the JxB line indicates that the effect is an isolated case.

Having earlier assessed the effect of time on allele-specific CRISPR/Cas9 mutagenesis, the design of the time course experiment for *Impact* (shown in **Figure 4.1**) also allowed assessing the effect of shorter periods of exposure to the allele-specific HDR frequencies. Thought this data is only available for one locus, it appears that shorter times of exposure to the genome editing reagents do not influence the effect of chromatin modifications on allele-specific HDR frequencies. As a reminder, this time course was performed with the px459v2 vector (Addgene Plasmid #62988) in the absence of puromycin selection due to the reduced time frame. A schematic for the harvesting of time points is provided in **Figure 5.2.A**. HDR frequencies are plotted separately for each allele over time in Figure **5.2.B**. The fold difference of paternal/maternal HDR is also shown across time in **Figure 5.2.C**. This time the frequency of HDR appeared to be subtly elevated for the maternal inaccessible allele at early time points, but this difference was not statistically significant, as determined by paired two sample t tests performed at each time point (**Figure 5.2.B**). Further statistical analysis with 1-way ANOVA also showed that there was no significant difference in the fold difference of allele-specific HDR frequencies between the different time points, as also implied by the standard deviation error bars in place for each time point in **Figure 5.2.C**. These error bars are large, especially for the earliest time points at 8 hours

and 12 hours, which could be owed to the very low overall mutagenesis frequency (as shown in **Figure 4.1.B**).

Combined, the data from all five sgRNAs, and the time course for *Impact*, suggest that pre-existing chromatin modifications on the target sites do not influence the frequency of HDR towards repair of Cas9-induced DSBs. HDR frequencies range from 10% to 40% for different target sites. This could be owed to the sequence-specific properties of each site and/or donor repair template used. Overall, accessibility and repressive chromatin modifications have little influence on the prevalence of HDR during genome editing with CRISPR/Cas9.

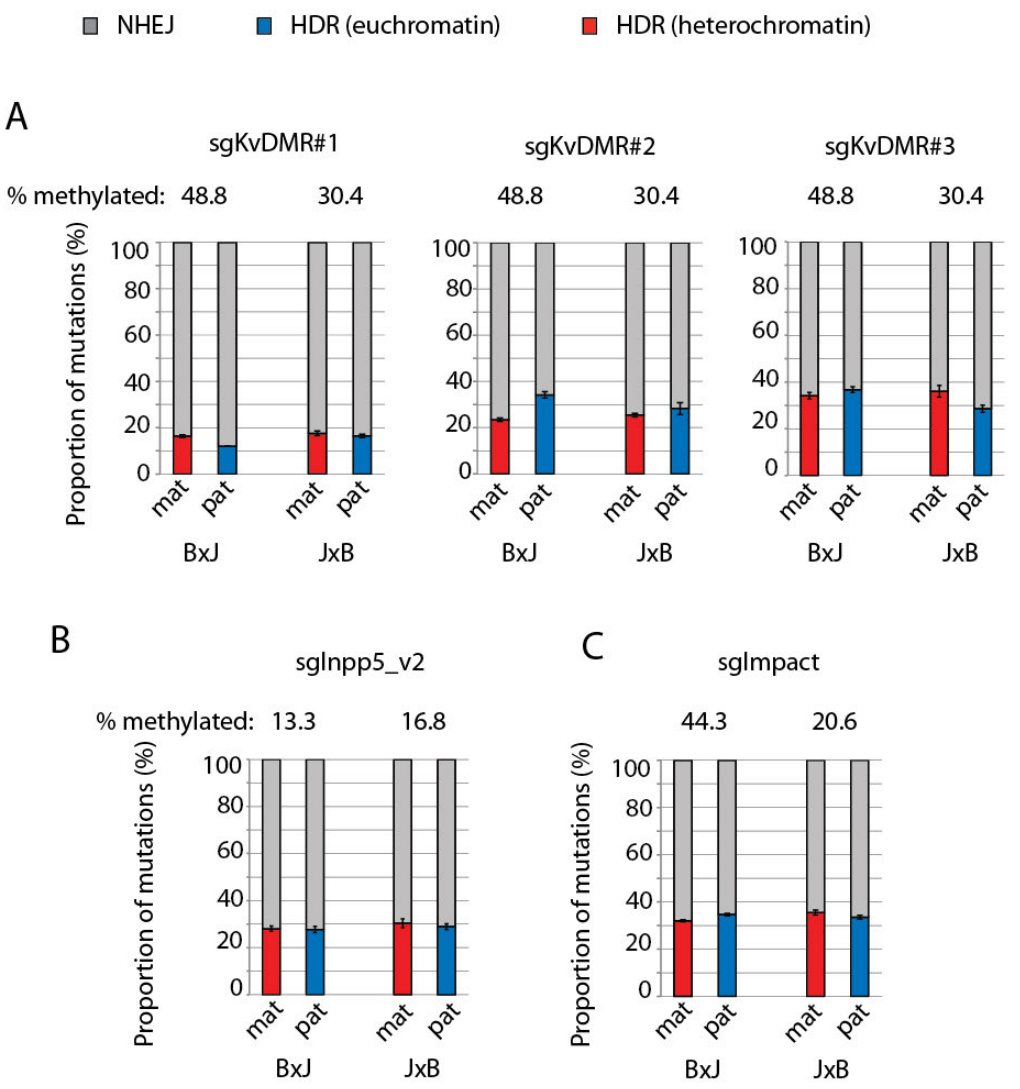


Figure 5.1 Chromatin modifications do not affect the frequency of repair via a single-stranded donor template. A: HDR frequencies for three individual sgRNAs on the *KvDMR* region, colour-coded in an allele-specific manner. Statistical analysis through paired two sample t tests from three experimental replicates revealed no significant difference between the maternal and paternal allele frequency. Bonferroni correction was performed to correct for type 1 error among these 10 individual experiments. The degree of imprinting maintenance is indicated by the percentage of fully methylated reads above each graph. The position of each sgRNA is indicated in **Figure 3.2**. **B:** HDR frequencies for the sgRNA on the promoter region of *Inpp5f_v2*, as for above. The position of sgInpp5f_v2 is indicated in **Figure 3.4**. **C:** HDR frequencies for the sgRNA on the promoter region of *Impact*, as for above. The position of sgImpact is indicated in **Figure 3.3**.

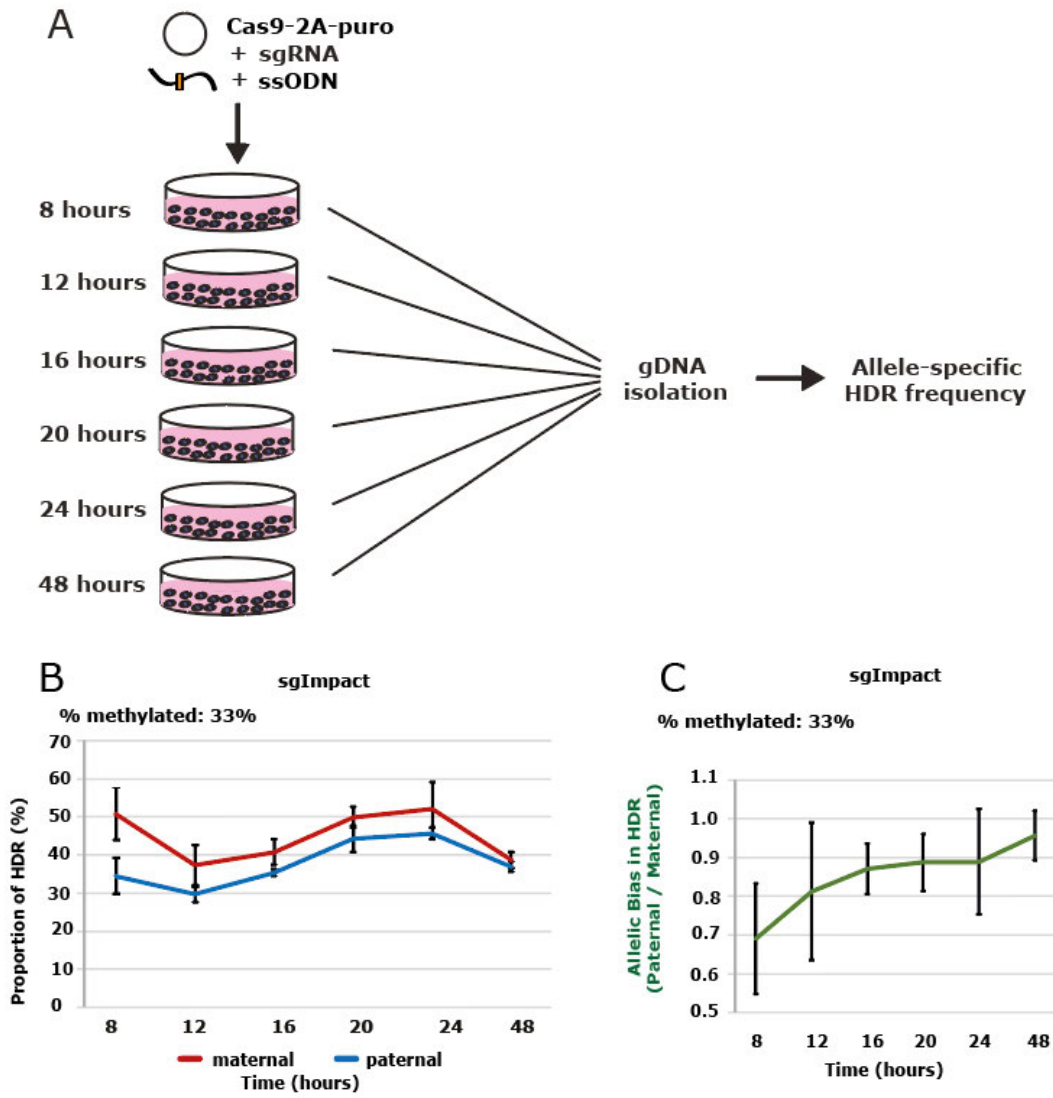


Figure 5.2 Chromatin modifications do not affect the frequency of HDR during CRISPR/Cas9 mutagenesis across time. A: Schematic of the experimental time frame for the time course experiment. BxJ cells were transfected px459v2 containing sgImpact and a ssODN. Cells were harvested at the indicated 4-hour intervals and genomic DNA was isolated for targeted amplicon NGS as previously described. **B:** Allele-specific HDR frequencies plotted against time. The paternal accessible allele is shown in blue and the maternal inaccessible allele is shown in red. HDR is slightly yet insignificantly more frequent on the maternal allele, as determined by paired two sample t tests per time point. The percentage fully methylated reads shown above implies the degree of imprinting maintenance during this experiment. **C:** Fold difference of HDR frequencies plotted against time. The green line connects the values per each time point. 1-way ANOVA and Tukey's HSD test showed the lack of a significant difference between every time point.

5.3 Homology-directed repair does not occur more readily within actively transcribed gene bodies

Having observed a lack of significant difference so far, I decided to investigate additional regions that possess epigenetic modifications associated with a higher prevalence of traditional HR. Since H3K9me3 and H4K20me3 are not constant across imprinted domains (Mikkelsen *et al.*, 2007), which include IRCs and imprinted genes, we sought to compare the effect of different classes of chromatin modification, such as H3K36me3 that had previously been implicated in HR efficiency. In 2014 Aymard *et al* established the presence of a link between H3K36me3 and the use of HR to repair DSBs that arise within actively transcribed gene bodies.

I selected to target *Impact* and *Grb10*, which are two maternally-imprinted loci that are highly expressed in mESCs (Maupetit-Méhouas *et al.*, 2016; Bouschet *et al.*, 2017). Our established system would allow us to investigate if the frequency of HDR would also be disproportionally elevated at the paternal transcribed allele of these two loci, as previously demonstrated for HR (Aymard *et al*., 2014). The sgRNAs for both loci were selected to target 3´ end distal introns, to avoid disturbing the expression of the two genes. In addition, the positions of the sgRNAs correlated with regions of high enrichment of H3K36me3 in mESCs, as shown in **Figure 5.3.A** for *Grb10* and **Figure 5.4.A** for *Impact*, respectively. The H3K36me3 ChIP-Seq tracks were obtained from publicly available ENCODE data through the UCSC genome browser. Although these ChIP-Seq data are from E14 mESCs cultured in serum+LIF, instead of 1i culture conditions, they are the closest publicly available data I could use to predict where H3K36me3 would be enriched in the two hybrid mESC lines I used.

Similar to the experiments described in the previous section for promoter-proximal regions, the data represented here was obtained at 4 days post transfection. It is evident that the HDR frequencies are very similar to the

values described previously, and no significant enrichment for HDR on the expressed paternal was observed. This is true for both the 3´ end of *Grb10* (**Figure 5.3.B**) and *Impact* (**Figure 5.4.B**) in both mESC lines used throughout my project.

However, these experiments were performed before I developed the strategy for monitoring imprinting maintenance that is described in **Figure 3.6**. Therefore, I have not reported on the degree of imprinting maintenance at the time and passage of cells were these editing experiments were conducted, as also mentioned in the figure legends of **Figure 5.3** and **Figure 5.4**. Moreover, native histone ChIP followed by qPCR has not been performed to assess the presence of H3K36me3 at the time. Nevertheless, it is possible that HDR is not influenced like HR does by H3K36me3, for reasons that will be listed in **Chapter 6**.

The overall conclusion that is drawn from the data for all seven different sites studied within and outwith maternally-imprinted gene bodies is that the HDR is employed just as frequently on either allele. Therefore, the pre-existing chromatin landscape does not influence the outcome of DSB repair during genome editing with CRISPR/Cas9.

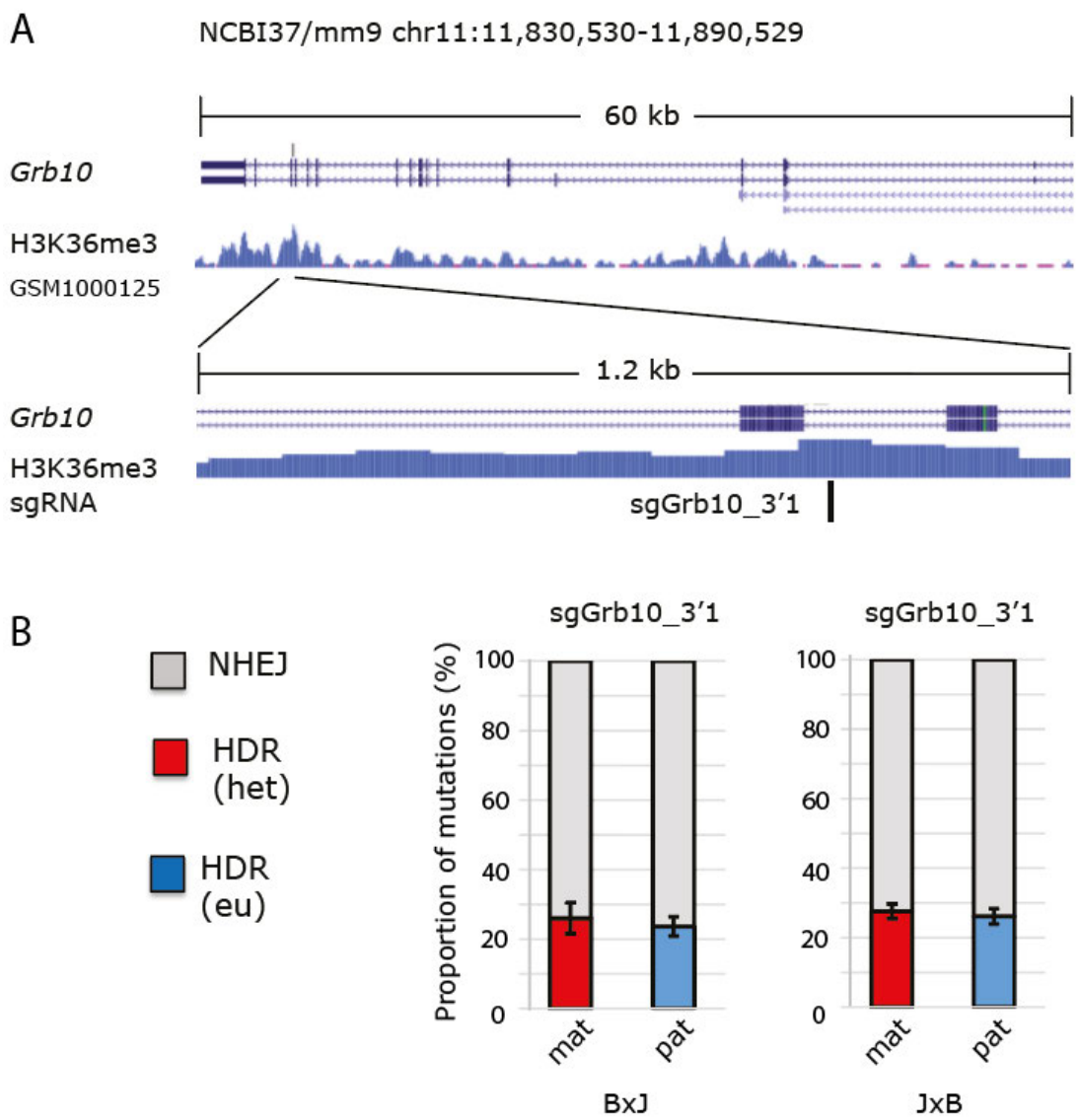


Figure 5.3 HDR frequency is not increased within the gene body of the actively transcribed allele of *Grb10*. A: View of the distal region of *Grb10* and the relative positions of sgGrb10_3´1. **B:** HDR frequencies for the sgRNA on the 3´ end distal intronic region of *Grb10*, colour-coded in an allele-specific manner. Statistical analysis through paired two sample t tests from three experimental replicates revealed no significant difference between the maternal and paternal allele frequencies for either hybrid mESC line. Bonferroni correction was performed to correct for type 1 error among the 4 individual experiments included in the present and next Figure. The degree of imprinting maintenance has not been determined for this locus at the time of targeting.

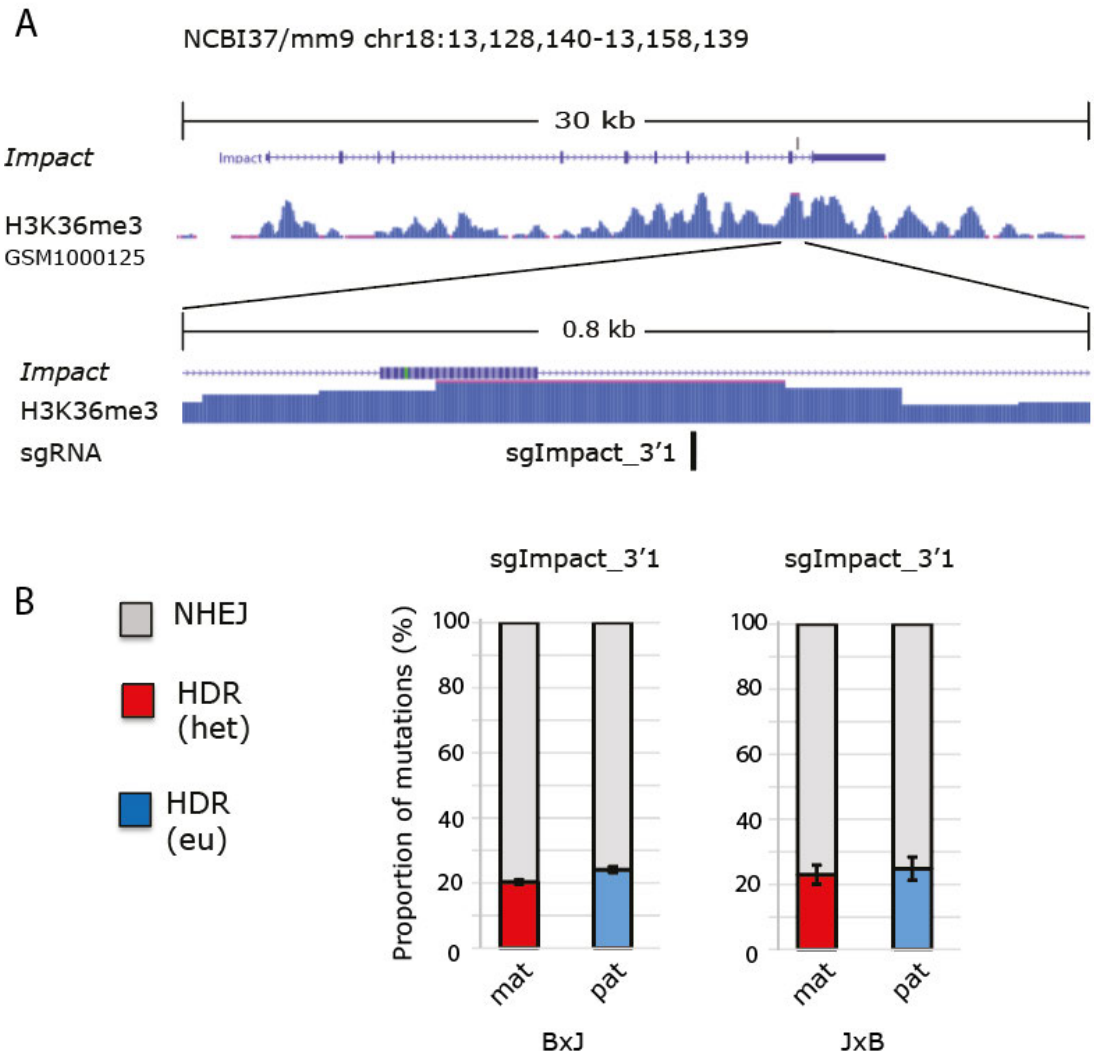


Figure 5.4 HDR frequency is not increased within the gene body of the actively transcribed allele of *Impact*. A: View of the distal region of *Impact* and the relative positions of sgImpact_3´1. **B:** HDR frequencies for the sgRNA on the 3´ end distal intronic region of *Impact*, colour-coded in an allele-specific manner. Statistical analysis through paired two sample t tests from three experimental replicates revealed no significant difference between the maternal and paternal allele frequency for only one of the two reciprocal hybrid mESC lines. Bonferroni correction was performed to correct for type 1 error among the 4 individual experiments included in the present and previous Figure. The degree of imprinting maintenance has not been determined for this locus at the time of targeting.

5.4 The InDel spectrum is not affected by chromatin during genome editing

In addition to the frequency of HDR, we also characterised the sizes and recurrence of different types of InDels produced by NHEJ in an allele-specific manner.

During a DNA repair profiling study the authors demonstrated that the repair of DSBs generated by the standard Cas9 endonuclease from *Streptococcus pyogenes* is influenced by the sequence of the protospacer instead of the wider genomic context (van Overbeek *et al.*, 2016). With the aid of Alison Meynert and Tracy Ballinger we focused on the sequencing reads that contained evidence of mutagenic NHEJ and assessed the frequencies of specific types of small deletions and single base insertions. All the results shown in **Figure 5.5** and **Figure 5.6** are also from edited genomic DNA samples collected 4 days post transfection.

Overall, it is evident that a unique pattern of InDels is generated after editing with every individual sgRNA. The majority of InDels found across all five studied target sites fell into distinct categories according to size. The five most frequent InDel classes for the five of the sgRNAs used throughout this study are shown in **Figure 5.5** and **Figure 5.6**. These five classes are different for each sgRNA. The variance across three experimental replicates is provided as standard deviation for each indel category of the sgRNAs in **Appendix 6**.

When sorted according to parental origin, the frequency for each InDel category was similar between the two alleles across three maternally-imprinted loci. This trend is observed for sgKvDMR#1 (**Figure 5.5.A**), sgKvDMR#2 (**Figure 5.5.B**), sgKvDMR#3 (**Figure 5.5.C**), sgInpp5fv_2 (**Figure 5.6.A**) and sgImpact (**Figure 5.6.B**). This suggests that the pre-existing chromatin modifications do not affect the outcome of mutagenic DNA repair during genome editing with CRISPR/Cas9.

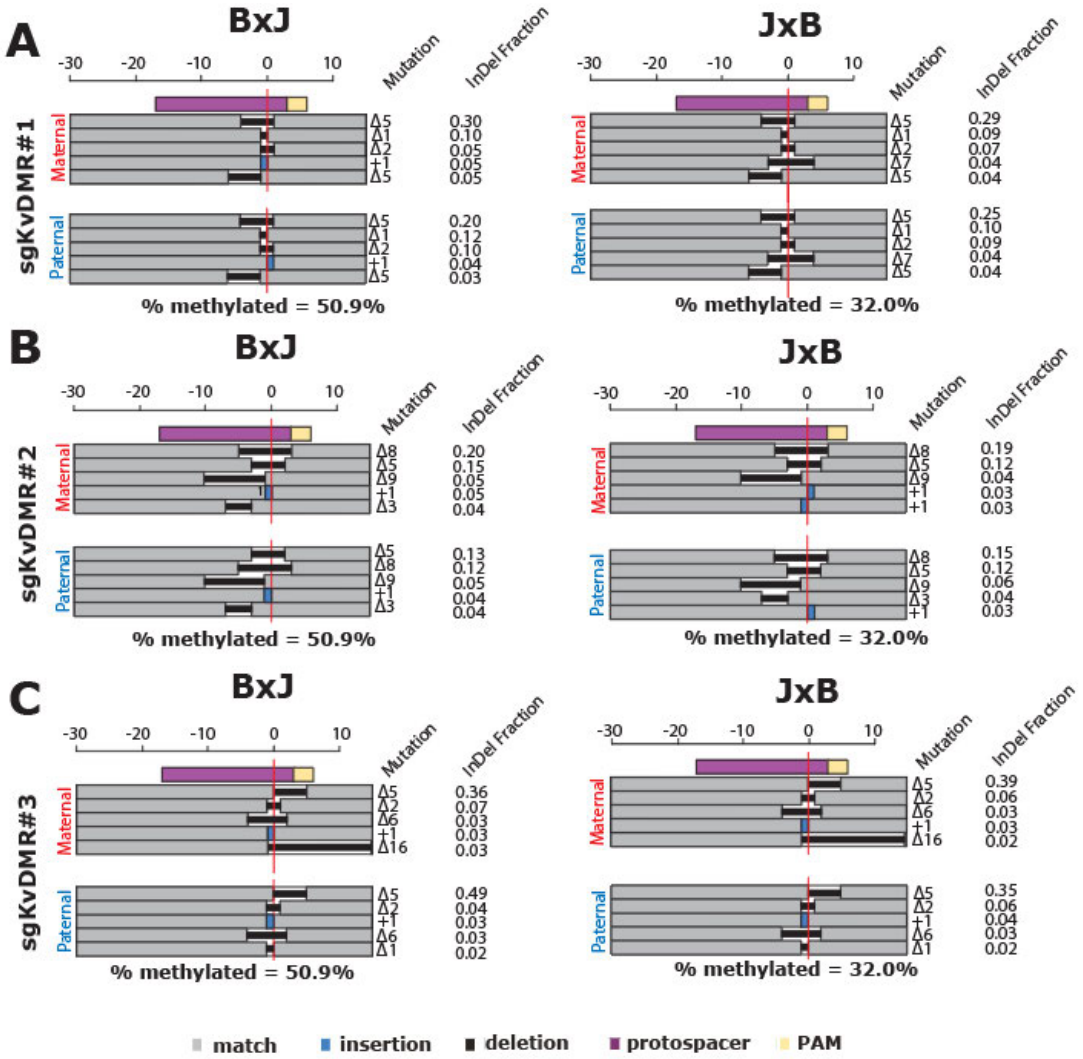


Figure 5.5 Recurring InDel classes at three sites on the *KvDMR* region.

A: Size and frequency of the most common InDels for sgKvDMR#1 for each allele. The position of the cleavage site is noted by the vertical line. Results from the BxJ and JxB lines are shown on the left side and right side, respectively. The scale above each graph illustrates the size of deletions, whereas the only single base insertions were detected. The degree of imprinting maintenance is indicated by the percentage of fully methylated reads is indicated below each graph. Colour coding is used for all graphs, with a key provided at the bottom. **B:** Size and frequency of the most common InDels for sgKvDMR#2 for each allele, as described above for sgKvDMR#1. **C:** Size and frequency of the most common InDels for sgKvDMR#3 for each allele, as described above for sgKvDMR#1. **Plots designed by Andrew Wood.**

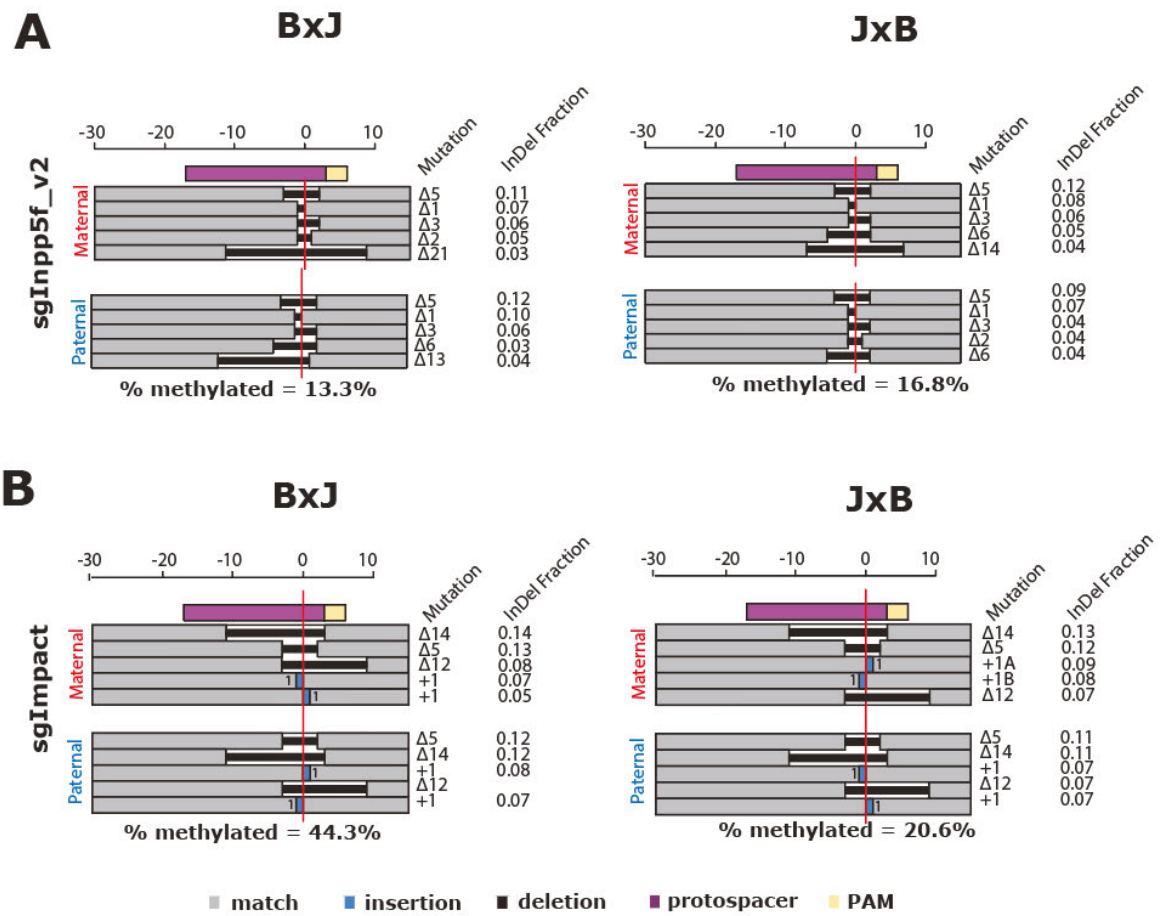


Figure 5.6 Recurring InDel classes at the sites on the promoter regions of *Inpp5f_v2* and *Impact*. A: Size and frequency of the most common InDels for sgInpp5f_v2 for each allele. The position of the cleavage site is noted by the vertical line. Results from the BxJ and JxB lines are shown on the left side and right side, respectively. The scale above each graph illustrates the size of deletions, whereas the only single base insertions were detected. The degree of imprinting maintenance is indicated by the percentage of fully methylated reads is indicated below each graph. Colour coding is used for all graphs, with a key provided at the bottom. **B:** Size and frequency of the most common InDels for sgImpact for each allele, as described above for sgInpp5f_v2. **Plots designed by Andrew Wood.**

5.5 Changes in the InDel profile of sgImpact over time

As previously demonstrated by van Overbeek *et al*, the relative frequency of different InDel classes is influenced over time, provided that Cas9 expression is continuously expressed (van Overbeek *et al.*, 2016). According to Kim *et al*, we assumed Cas9 expression to be maintained at high levels for at least 72 hours from plasmid DNA we used (instead of RNPs) to deliver CRISPR/cas9 into mESCs (Kim *et al.*, 2014). While performing the time course experiment on the promoter region of *Impact* (as previously described in **Chapter 4**, **Figure 4.1**) ssODN was co-delivered with CRISPR/Cas9 for us to assess the effect of chromatin on repair of Cas9-induced DSBs over a 48-hour timeframe. Alison Meynert performed the necessary analysis to distinguish single nucleotide deletions into a class (Δ1). She also grouped 10 nucleotide-long deletions with deletions that were larger than 10 nucleotides in length into a single class (Δ⩾10). The frequency of both classes was plotted against time and shown separately for the euchromatic and heterochromatic allele in **Figure 5.7**.

As shown in **Figure 5.7.A**, the frequency of Δ1 has a tendency to decrease over time for either allele. On the contrary, the frequency of class Δ≥10 showed an increasing trend over time (**Figure 5.7.B**). These findings agree with published results from van Overbeek *et al*, and may be explained by the fact that small deletions are less likely to perturb the PAM site and protospacer sequence than larger deletions. Integrity of the target site may result in repeated cycles of Cas9 cleavage and the generation of larger deletions, such as Δ⩾10.

Despite observing that the frequency of class Δ⩾10 increases in both the accessible and imprinted allele for this locus (**Figure 5.7.B**), we also observed that the frequency of class Δ1 was higher on the imprinted maternal allele for the first 24 hours (**Figure 5.7.A**). According to **Figure 4.1.B**, the overall editing frequency is lower for the imprinted allele, compared to the

accessible paternal allele, for these early time points. This may suggest that repetitive cycles of cleavage by Cas9 do not occur as readily on the inaccessible, imprinted maternal allele. Although these findings show a logical trend, it would be worth to perform additional time course experiments in the future to assess more imprinted target sites in this manner.

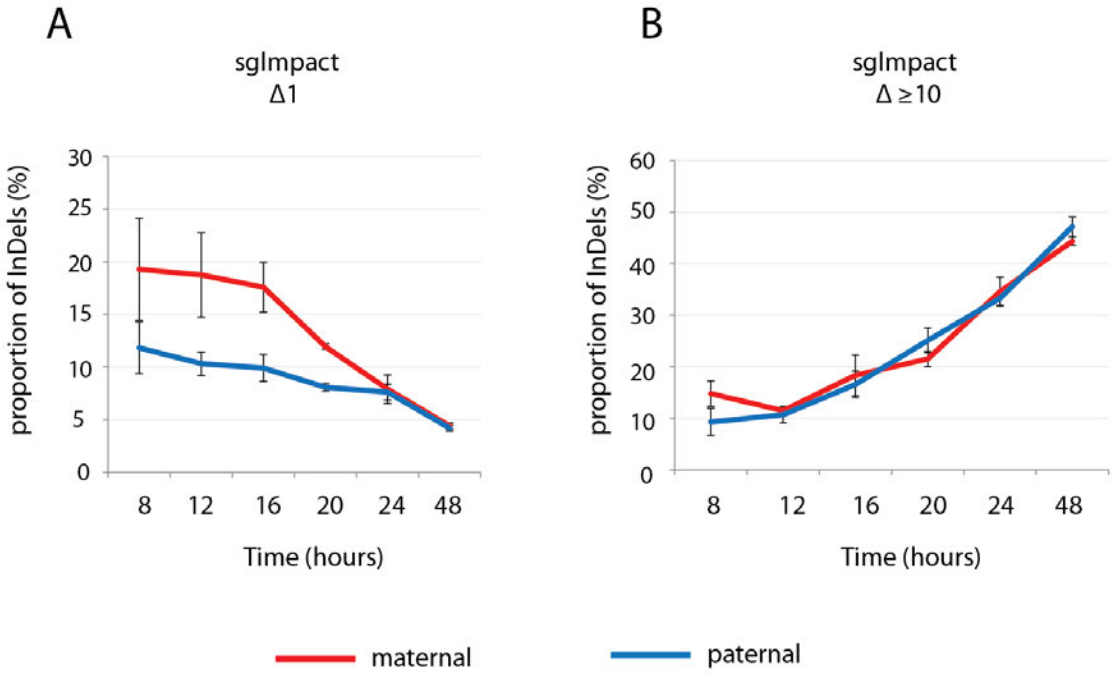


Figure 5.7 Larger deletions accumulate over time on the promoter of *Impact*.

A: Allele-specific changes in the proportion of reads that contain a single nucleotide deletion **(Δ1)**. The frequency of this class of (small) deletion is reduced over time for both alleles. The data was collected from the 48-hour time course experiment previously shown in **Figure 4.1** for sgImpact. Error bars represent Standard Deviation from 3 experimental replicates. **B:** Allele-specific changes in the proportion of reads that contain deletions of 10 nucleotides or more **(Δ≥10)**. The frequency of this class of (larger) deletion increases over time for both alleles. The data was collected from the 48-hour time course experiment previously shown in **Figure 4.1** for sgImpact. Error bars represent Standard Deviation from 3 experimental replicates.

5.6 Discussion

Despite evidence that suggests DNA repair is influenced by the sequence context and the position of a DSB within the nucleus (Lemaître *et al*., 2014; Aymard *et al*., 2014; Tsouroula *et al*., 2016) we discovered that differential chromatin modifications between two alleles of three distinct maternally-imprinted loci do not have an effect to the selection between NHEJ and HDR towards the repair of Cas9-induced DSBs.

As previously suggested by two different research groups, Cas9 may open up compacted chromatin regions (Polstein *et al*., 2015; Barkal *et al*., 2016). This might explain why pre-existing chromatin modifications do not contribute to DNA repair pathway selection during genome editing. Our observations are in agreement with a recent study, where DSB repair pathway choice was not affected by chromatin compaction in *Drosophila melanogaster* after DSBs were induced with the I-SceI endonuclease (Janssen *et al.*, 2016).

Regarding the InDel analysis performed for five sgRNAs, we observed that the InDel spectrum for every unique sgRNA was different. This became evident after recording the five most frequent InDel classes for each sgRNA. As mentioned earlier, this agrees with the finding that the sequence of the protospacer region is responsible for the produced InDel pattern, published by van Overbeek *et al.* in 2016. Emerging evidence continues to support the idea that the DNA sequence context of the target site can be used to predict the repair outcome of the Cas9-induced DSB in the absence of a repair donor either *in vitro* (Allen *et al.*, 2018; Shou *et al.*, 2018) or *in vivo* (Shen *et al.*, 2018). However, our findings suggest that differential chromatin modifications do not influence the repair outcome of Cas9-induced DSBs when the latter are repaired without relying on an externally provided repair donor.

Our system allows us to quantify mutagenesis by CRISPR/Cas9 by measuring the frequency of reads that deviate from the wild-type unedited sequence of the murine genome. Therefore, we are only able to detect

mutagenic DNA repair and cannot account for the frequency of error-free c-NHEJ, during which there is no InDel formation. Nevertheless, we report that genome edits arising via HDR and NHEJ occur at similar frequencies for DSBs generated by Cas9 on transcriptionally silenced and transcriptionally active genomic sites. All findings from this chapter will be discussed further in the upcoming discussion chapter.

Chapter 6

Discussion and Future Directions

6.1 General Discussion

The development of the CRISPR/Cas9 genome editing tool is undoubtedly one of the greatest recent advances in molecular genetics. CRISPR/Cas9 is broadly adaptable and will continue to revolutionise the way biological research is conducted in the future. The natural origin of the standard version of this tool lies in the Type II CRISPR bacterial immunity system from *Streptococcus pyogenes*, where an RNA-guided endonuclease degrades invading DNA molecules following RNA-DNA target hybridisation.

The genomes of higher eukaryotes are organised into chromatin, a higher order structure that compacts genomic DNA to fit into the nucleus. Despite being widely used, CRISPR/Cas9 does not naturally recognise chromatinised DNA as a substrate. Indeed, several studies have shown that nucleosomes protect DNA from cleavage by CRISPR/Cas9 *in vitro* (Hinz, Laughery and Wyrick, 2015; Horlbeck *et al*., 2016; Isaac *et al*., 2016). Studies using cell-based systems have also shown that chromatin compaction reduces the efficiency of CRISPR/Cas9 mutagenesis (Chen *et al.*, 2016; Fujita, Yuno and Fujii, 2016; Daer *et al.*, 2017). However, Cas9 can penetrate highly compacted heterochromatic regions and has been proposed to remodel nucleosomes inside mammalian cells (Knight *et al.*, 2015; Polstein *et al.*, 2015; Barkal *et al.*, 2016). This creates a paradoxical situation that merits further investigation.

In the past, the use of genomic imprinting as a model system has led to key discoveries about the regulation of gene expression by factors like CTCF and lncRNAs, as well as genome organisation (Hark *et al.*, 2000; Murrell, Heeson and Reik, 2004; Nagano *et al.*, 2008). Imprinted genes must only be expressed from one of the two inherited parental alleles in mammals, in order to secure proper prenatal development. This is achieved by the establishment of ‘imprints’, which consist of differential deposition of DNA methylation on the two alleles. This event is followed by additional deposition of repressive

histone modifications, such as H3K9me3 and H4K20me3, on the allele that needs to be silenced and results in the establishment of a small-scale heterochromatic environment that is frequently embedded within wider euchromatic regions (Regha *et al.*, 2007).

The presence of the same underlying DNA sequence in two different chromatin conformations within the same nuclear environment created an excellent opportunity for us to investigate the effect of chromatin on CRISPR/Cas9 editing efficiency. By repurposing genomic imprinting, we developed a system to quantify mutagenesis in an allele-specific manner after co-delivery of CRISPR/Cas9 and an ssODN into two reciprocal hybrid mESC lines. The presence of repressive chromatin modifications on one of the two parental alleles was verified in both mESC lines and the degree of imprinting maintenance was monitored throughout every targeting experiment.

A major advantage of this system is that a repressed and an accessible allele are present at a one-to-one ratio inside each cell nucleus, providing an ideal internal control. This was a very big caveat of other cell-based systems that have been used so far, which relied on perturbing the chromatin environment where a reporter gene was located (Chen *et* al., 2016; Daer *et al*., 2017). Our system avoids perturbing the natural configuration of chromatin within mammalian cells, because genomic imprinting is a natural phenomenon. Furthermore, both of the studies mentioned above have been transfecting separate cell populations (untreated vs doxycycline-treated) to infer about the effect of chromatin on the editing efficiency of CRISPR/Cas9. As we demonstrated in **Figure 4.3**, plasmid transfections lead to a heterogeneous cell population, where different Cas9 concentrations affected the impact of chromatin on the editing efficiency of CRISPR/Cas9. Neither Chen *et al* nor Daer *et al* controlled for the concentration of Cas9 between the different transfections they were performing on the separate cell populations that were untreated or treated with doxycycline. This might explain some of the enigmatic higher magnitudes of chromatin-related effects they reported, compared to our earliest findings (**Figure 3.7**).

Our earliest editing experiments on three separate maternally-imprinted loci proved the presence of a subtle yet reproducible bias for editing on the paternal non-imprinted allele following exposure to the genome editing reagents for 4 days (**Figure 3.7 and Figure 3.8**). The puromycin selection that was performed during these experiments may enrich for the cells with very high Cas9 expression from the introduced plasmid. As shown in **Figure 4.3** and **Figure 4.4**, high Cas9 expression results in lower allelic bias and provides more opportunity for the less frequently accessed allele to be cleaved and edited. However, similar subtle allele-specific bias values are also observed in the absence of puromycin selection during the 4 day-long exposure experiments shown in **Figure 4.2**.

Despite observing lower bias levels between the euchromatic and heterochromatic regions we assessed, compared to the bias levels observed by the studies mentioned above, we were able to vary and test other parameters that could influence the way chromatin compaction affects CRISPR/Cas9 mutagenesis. This is owed to the natural, internally controlled system we employed, where two chromatin states were assessed in parallel within the same population of cells. Having observed differences in the magnitude of the effect of chromatin between separate transfections, or between different loci, western blots could have been performed to report on the relevant expression levels of Cas9 within cells following transfection. A serious caveat of our system was the partial loss of imprinting we frequently experienced as a consequence of maintaining the two reciprocal mESC lines in 1i culture conditions (Choi *et al.*, 2017). Constant monitoring of the degree of imprinting maintenance via targeted sequencing of bisulfite amplicons helped us determine the impact of imprinting loss on each experimental dataset.

While our work was ongoing, a study using a different internally-controlled system emerged. This study assessed a single locus in the HCT116 human colon cancer cell line, and primarily focused on the effect of DNA methylation on CRISPR/Cas9 editing efficiency (Fujita, Yuno and Fujii, 2016). Three different sgRNAs were used to edit the *p16INK4a* locus, in separate

transfections for every sgRNA. The authors concluded that the editing efficiency of only one of those three sgRNAs was affected by chromatin modification, as inferred from differential DNA methylation. We also used three different sgRNAs to target the *KvDMR* locus. As shown in **Figure 3.7**, we observed that chromatin modifications, including repressive histone modifications and DNA methylation, reduced the efficiency of all three sgRNAs that were used to target the *KvDMR* locus. This could be attributed to the fact that we used a high-throughput assay involving Next generation sequencing to quantify the effect of chromatin on CRISPR/Cas9 mutagenesis instead of cloning PCR amplicons from edited genomic DNA samples into plasmids and performing Sanger sequencing to obtain a smaller number of reads.

By reducing the editing timeframe and sampling targeted cells at 4-hour intervals, as shown in **Figure 4.1**, we discovered that the impact of repressive chromatin modifications is more profound during short timeframes of exposure to the genome editing reagents. This supports previous findings showing that heterochromatin delays the search of CRISPR/Cas9 for its target site (Knight *et al*., 2015). In addition, we showed that the lower intracellular concentration of eGFP, and therefore Cas9, the higher the ablation of allelic bias towards the accessible non-imprinted allele (**Figure 4.3**). Finally, the presence of designer mismatches between the sgRNA and the target site resulted in a disproportionate reduction of mutagenesis on the imprinted heterochromatic allele, implying that chromatin compaction reduces off-target CRISPR/Cas9 mutagenesis (**Chapter 4.5**). This finding is consistent with a recently published study where differentially accessible genomic regions were targeted in HEK293 and HeLa cells (Kim and Kim, 2018). The system employed by Kim and Kim involved using mismatch-containing sgRNAs to target two sequences, similar to the two loci we targeted, that were present in two copies in the human genome. However, the two target sites were selected based on DNase hypersensitivity alone in the absence of other epigenetic modification differences between the two copies of the target sequences. Our system is based on differential accessibility to DNase I, differential DNA methylation and

repressive histone modifications, which are overall more informative on the effect of natural chromatin compaction than DNase I accessibility alone.

Additionally, we showed that dCas9 associates more readily with the accessible, non-imprinted allele between 8 and 96 hours after lipofection, which agrees with conclusions from previous genome-wide ChIP-seq studies for global assessment of dCas9 occupancy (Kuscu *et al.*, 2014; Wu *et al.*, 2014). Except for the possibility that Cas9 can remodel chromatin, it is also likely that Cas9 gains access to genomic sites that are occluded by pre-existing repressive chromatin modifications as a result of opportunistic nucleosome remodelling. Taken together, our data provide novel insights into how chromatin regulates CRISPR/Cas9 mutagenesis and extend the findings of previous studies on this subject.

Looking forward, chromatin compaction should be considered when designing genome editing experiments and selecting sgRNAs with as few predicted off-target site effects as possible. In general, it is desirable to pick sgRNAs that have a high on-target efficiency and a low off-target efficiency. By combining these two parameters we can derive the ratio of **on-target/off-target** mutagenesis. This ratio will be high if the on-target efficiency is also high and the off-target efficiency is low, as is preferable while designing editing experiments.

Our data suggests that while the intracellular concentration of Cas9 is low, and Cas9 exposure is brief, chromatin exerts the biggest effect at the on-target mutagenesis frequency. In this scenario, accessible euchromatic regions are edited much more readily than inaccessible heterochromatic regions. Regardless of the exposure time and intracellular Cas9 concentration, we and others (Kim and Kim, 2018) have observed that off-target mutagenesis is greatly reduced in inaccessible heterochromatic regions. Our data is thus adequate for us to propose a model for obtaining high **on-target/off-target** mutagenesis ratios while designing editing experiments with CRISPR/Cas9. In order to maximise on-target and minimise off-target mutagenesis, researchers should select sgRNAs whose target sites are located within euchromatin and

predicted off-target sites are located in heterochromatin, in any cell line of choice. To further refine the efficiency of on-target mutagenesis, exposure to the genome editing reagents should be brief and the intracellular concentration of Cas9 must be maintained at low levels. Our model is illustrated in detail in **Figure 6.1**.

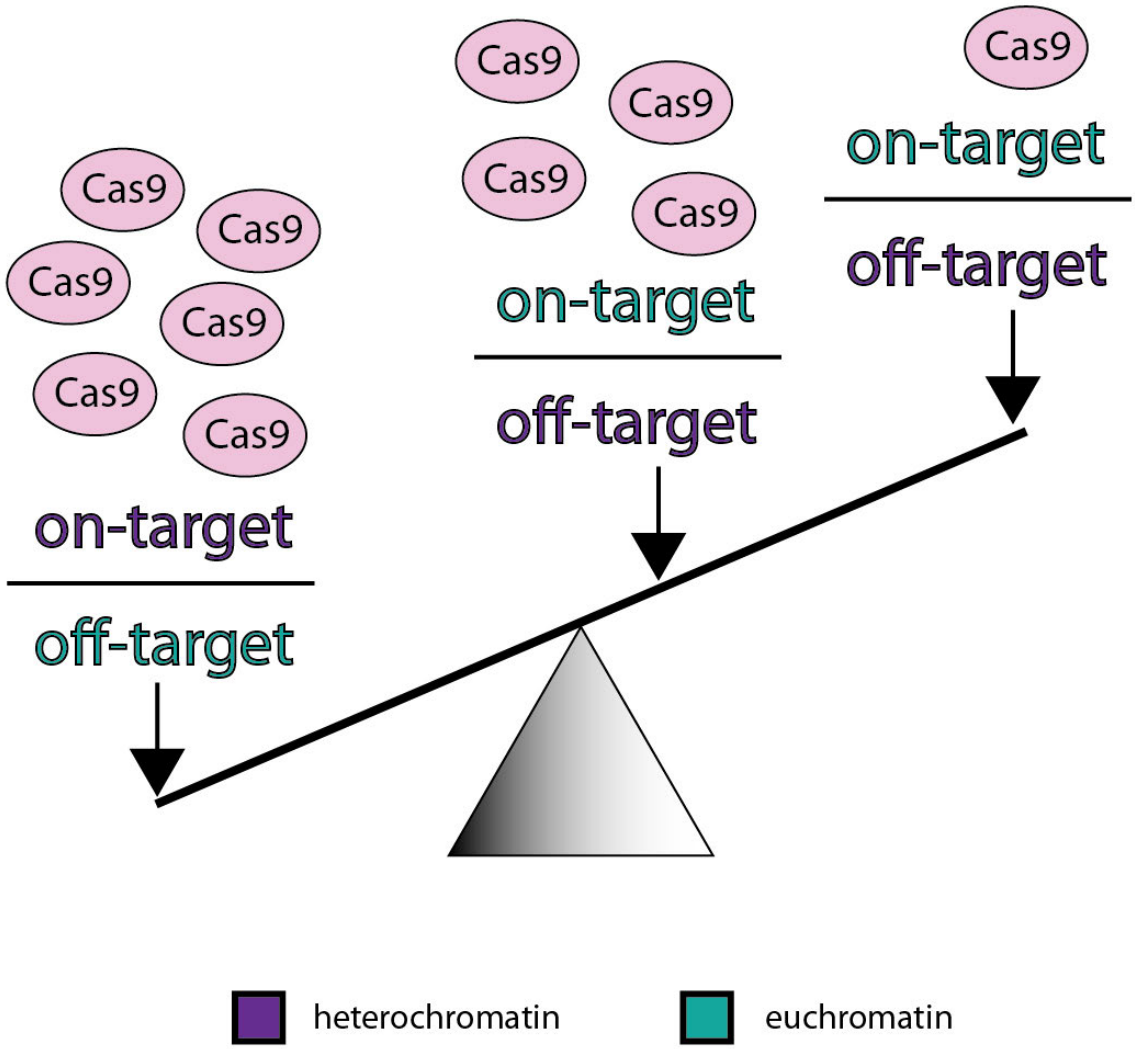


Figure 6.1 Model for acquiring a high on-target/off-target mutagenesis ratio.

Off-target site mutagenesis can be minimised by selecting sgRNAs that have predicted off-target sites in heterochromatin (represented by purple letters). In addition, on-target efficiency can be maximised by selecting sgRNAs that target euchromatic regions (represented by turquoise letters) while limiting the concentration of Cas9 during the experiment (as implied by the amount of Cas9 molecules, shown in pink). If all of these conditions are met, the **on-target/off-target** mutagenesis ratio is maximised (as shown on the right). If the concentration of Cas9 increases the on-target efficiency for euchromatic sgRNAs decreases, leading to a lower **on-target/off-target** mutagenesis ratio (as shown in the middle). Finally, the lowest possible **on-target/off-target** mutagenesis ratio is obtained by combining a high intracellular Cas9 concentration while using heterochromatic sgRNAs that have predicted off-targets within euchromatin (as shown on the left).

6.2 Can the effect of heterochromatin be refined in non-dividing cells?

Heterochromatin may generally be less permissive for transcription than euchromatin, but DNA replication must still occur during each cell cycle (Craig and Bickmore, 1993). The passage of the replication fork requires chromatin to be remodelled, which may lead to opportunistic binding of Cas9. It is also known that due to their special cell cycle arrangements, which are a very short G1 phase and extended S phase (White and Dalton, 2005), mESCs have a looser chromatin structure compared to differentiated cells (Meshorer, 2007), such as neural progenitors and mature neurons (Bonev *et al.*, 2017). Therefore, it is worth investigating what the effect of chromatin on CRISPR/Cas9 mutagenesis is by using other cell lines with better-defined heterochromatic regions.

Moving forward, I would like to target maternally-imprinted loci in post-mitotic cells and determine the effect of chromatin on targeted mutagenesis by CRISPR/Cas9 in the absence of DNA replication. If the effect we observe in post-mitotic cells is ablated compared to what we have observed in mESCs so far, we would suggest that chromatin exerts a larger effect in these cells. To explore this idea, we plan to target non-dividing hepatocytes in hybrid mice derived from the same reciprocal strain crosses as for our hybrid mESC lines. An increasingly popular method for direct delivery of the genome editing reagents to liver tissue is direct tail vein injection (Yang *et al.*, 2016), which has been recently optimised for delivery of naked plasmid DNA by our collaborator Dr Luke Boulter.

In the future, we would deliver our CRISPR-eGFP px458 constructs via tail vein injection to hybrid mice and determine the uptake efficiency of the reagents with Immunohistochemistry for eGFP. This approach will allow us to assess allele-specific CRISPR/Cas9 mutagenesis in the absence of DNA replication.

6.3 How is off-target site mutagenesis inhibited in heterochromatin?

Off-target site recognition is undesirable during genome editing with programmable nucleases. Despite the existence of several tools that predict off-target mutagenesis sites for sgRNA candidates, many of these tools ignore the native chromatin context of mammalian genomic DNA. During my PhD I introduced mismatches into the sgRNA, while still targeting the same maternally-imprinted loci, in order to mimic an off-target site and maintain the advantages of our internally controlled system. The early experiments from two different maternally-imprinted loci suggested that the native chromatin context impairs off-target site mutagenesis.

Our main focus for the future, will be to dissect the mechanism by which heterochromatin inhibits off-target mutagenesis by CRISPR/Cas9. In order to examine if heterochromatin impairs the initial association between the Cas9-sgRNA complex and its target site, we would like to use sgRNAs with mismatches on the protospacer sequence of the sgRNA molecule and perform ChIP followed by target amplicon sequencing. Similar to our approach for perfectly complementary sgRNAs, we would assess the degree of dCas9 binding in the absence of complementarity between the protospacer and target site in an allele-specific manner.

Furthermore, we would also like to perform additional experiments to identify which feature of heterochromatin is primarily responsible for this effect. The first heterochromatic feature in question will be DNA methylation, which does not influence the initial association of dCas9 *in vitro* (Fujita, Yuno and Fujii, 2016). To address if DNA methylation inhibits some of the later steps involved in CRISPR/Cas9 target recognition, such as R-loop formation, we would use an *in vitro* cleavage assay that compares the efficiency of CRISPR/Cas9 when the target DNA is either unmethylated or methylated. A diagram explaining the assay is provided in **Figure 6.2.A.**

The second feature of heterochromatin we are considering is nucleosome mobility. During the last months of my PhD I briefly undertook the design and cloning of DNA segments containing sgRNA into two separate and unique chromatin fibre constructs developed by Professor Nick Gilbert. The key difference between the two fibres lies in the affinity for nucleosome positioning affinity they possess. While the first fibre construct consists of a low affinity nucleosome binding site that is flanked by twelve repeats of the Widom 601 high affinity sequence, which result in a very well structured and stiff fibre (Lowary and Widom, 1998), the second fibre construct consists solely of low affinity nucleosome binding sites. An illustration of the nucleosome binding sites of the two fibre constructs is provided in **Figure 6.2.B**. The two constructs can be assembled into chromatin fibres with the aid of purified histone octamers from chicken erythrocytes (Feng, Scherl and Widom, 1993). Our goal is to compare how readily each of the two fibres are digested by purified Cas9 protein coupled with *in vitro*-transcribed sgRNA molecules in the presence and absence of mismatches on the protospacer. This will provide new insights into the effect of nucleosome positioning on off-target site cleavage and help us understand if nucleosome mobility, which is typically lower in transcriptionally repressed heterochromatin, could contribute to the disproportionate effect of mismatches on mutagenesis in this chromatin environment.

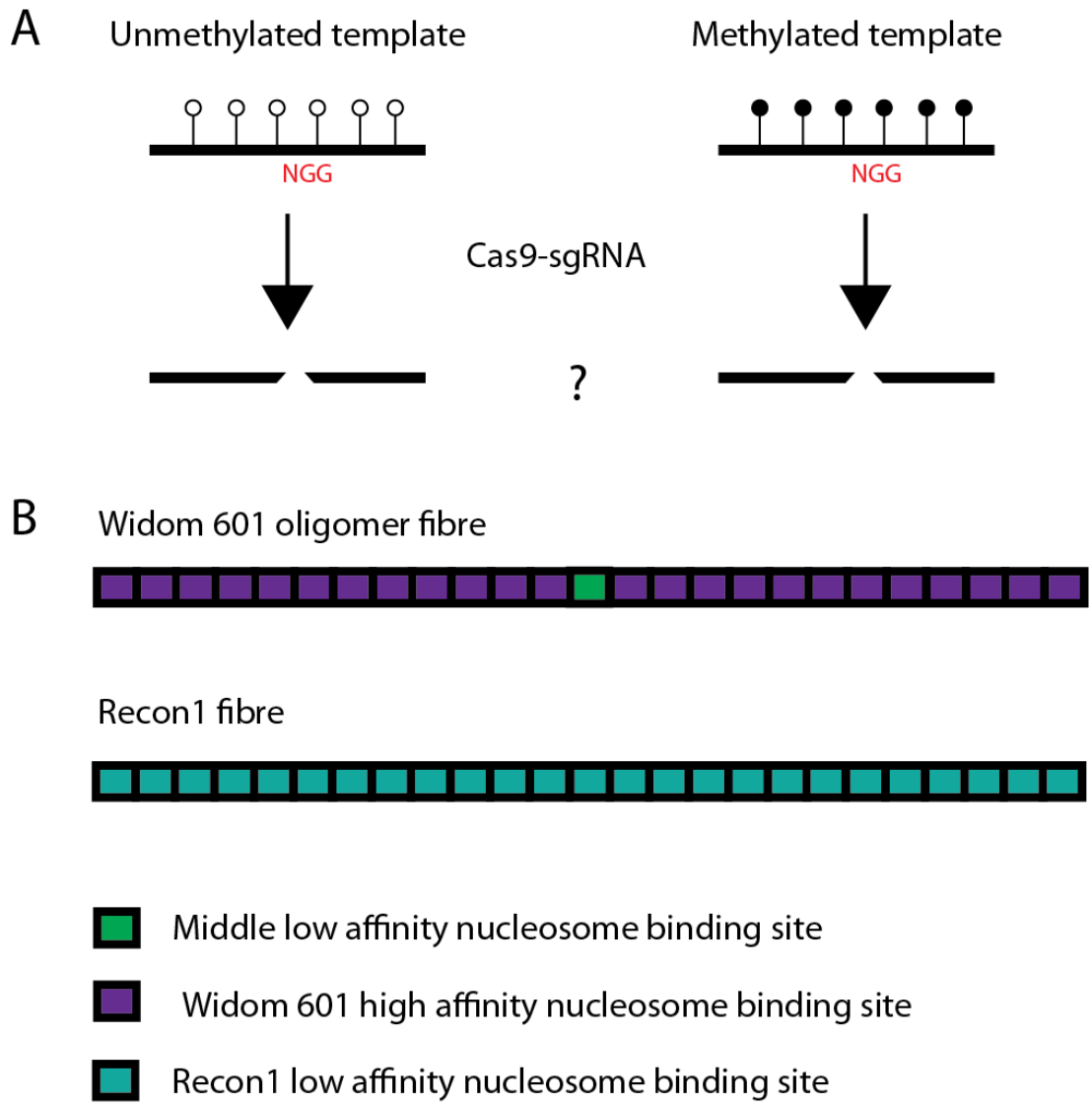


Figure 6.2 Schematic of the *in vitro* cleavage assay and the two chromatin fibre constructs. A: *In vitro* assay to assess and compare cleavage of unmethylated and methylated templates. The template sequence contains the PAM and protospacer for the previously used sgRNAs. Template methylation is also performed *in vitro* with M.SssI. Cleavage products will be visualised by agarose gel electrophoresis. **B:** Representation of the stiff Widom 601 oligomer fibre (top) and the Recon1 fibre (bottom). For the 601 Widom oligomer fibre, twelve high affinity 601 sites, shown in purple, are located on each site of a central low affinity nucleosome binding site, shown in green. The Recon1 fibre construct consists only of low nucleosome binding sites, shown in turquoise, and is less stiff than the 601 oligomer fibre. The PAM will be positioned at the centre of each fibre, on the central low affinity binding sites.

6.4 Pre-existing chromatin modifications do not influence Cas9-induced DSB repair

By leveraging the genomic imprinting-based system we developed, we also assessed the effect of chromatin on repair of Cas9-induced DSBs in the context of genome editing with CRISPR/Cas9. After assigning the parental origin to every generated sequencing read, all reads were broken down to three categories for each allele. Reads with small gaps or one extra nucleotide (InDels) were assigned to the NHEJ category, whereas reads containing the pre-engineered designer mutation from the externally provided single-stranded repair template were assigned to the HDR category. The reads where the sequence was identical the reference C57BL/6J (B) or JF1 (J) genomes were assigned to the third category, which was reads without evidence of CRISPR/Cas9 mutagenesis.

Allele-specific analysis of the reads bearing NHEJ-derived InDels revealed that every individual target site produced a unique InDel pattern that was not substantially influenced by chromatin state. This finding is relatable to two InDel profiling studies that did not asses the effect of chromatin and were published at different times throughout the course of my PhD (van Overbeek *et al.*, 2016; Chen *et al.*, 2019). During another recent study, that largely resembled that by van Overbeek *et al*, InDel profiles were assessed before and after treating HepG2 human liver cancer cells with Trichostatin A (TSA), which is a histone deacetylase inhibitor (Chakrabarti *et al.*, 2018). By altering the natural configuration of chromatin, the authors compared InDel profiles for compacted and de-compacted (after TSA treatment) genomic locations, using separate cell populations for each condition. The conclusion of this study was that the InDel profile of some sites, which do not consistently produce certain types of InDels, is affected by chromatin whereas sites that consistently produce the same types of InDels are not affected by changes in chromatin configuration (Chakrabarti *et al*., 2018). However, perturbing chromatin with

TSA in a cancer cell line is perhaps not the best way to derive conclusions about features that are shared across cell types, as suggested by the authors.

An important caveat of our assay relating to DNA repair is that the readout depends on mutation, and thus does not report on error-free c-NHEJ. However, it allowed us to fulfil our primary objective: to assess the effect of chromatin on genome editing. In the absence of resection or the lack of 1bp InDel generation during c-NHEJ, the endpoint after Cas9-induced DSB repair will be identical to the original target sequence. It is also possible to receive an output that is identical to the starting sequence if HR takes place from a sister chromatid, which is likely in mESCs due their unusually longer S-phase and shorter G1-phase (White and Dalton, 2005). Since these events would not produce a mutational consequence, they would be assigned into the spectrum of reads without evidence of mutation through our sequencing analysis pipeline. It is not currently possible to estimate what fraction of these 'non-mutagenised' reads corresponds to error-free c-NHEJ, HR from a sister chromatid or unedited strands, but our assay still reports on the most common outcomes of Cas9-induced DSB repair options (error-prone c-NHEJ, MMEJ, HDR). Overall, this caveat is shared between our assay and the other assays that have been used to assess the effect of chromatin on the outcome of CRISPR/Cas9 mutagenesis (Fujita, Yuno and Fujii, 2016; Daer *et al.*, 2017).

What remains unclear due to the small-scale range we assess with our targeted 150bp paired amplicon sequencing, is the occurrence of larger deletions through alt-EJ. Ultimately, we hope to subject the plethora of samples we generated during this project to an alternative sequencing method and generate longer sequencing reads. These will allow us to discover if there is a link between pre-existing chromatin modifications and larger deletions during genome editing with CRISPR/Cas9.

Concerning HDR, we discovered that it occurs at similar frequencies regardless of pre-existing chromatin modifications due to lack of statistically significant differences between the two alleles of the targeted maternally-imprinted genes. Additional evidence from a very recent study agrees with our

observations about HDR being unaffected by the pre-existing chromatin context of the target site (Janssen *et al.*, 2019). Another study assessing repair of DSBs induced by the I-SceI meganuclease in *Drosophila melanogaster* also reached a similar conclusion about the effect of chromatin compaction on HR occurrence (Janssen *et al.*, 2016). Although at first thought this strengthens our findings, it should be noted that instead of traditional HR we are assessing HDR through externally provided ssODNs, a mechanism that is still not fully characterised.

Our primary hypothesis was that HDR would more prevalent in the accessible euchromatic allele in our genomic imprinting-based system. The foundations of this hypothesis lie on a study that showed that HR is preferentially utilised to repair DSBs occurring within actively transcribed regions of the genome (Aymard *et al.*, 2014). According to Aymard *et al*, LEDGF mediates recruitment of CtIP to DSBs in actively transcribed gene bodies through direct association with H3K36me3. Although neither allele of imprinted promoters is decorated with H3K36me3, as shown in **Figure 5.3** and **Figure 5.4** the Cas9-induced DSBs we generated within actively transcribed imprinted gene bodies were not preferentially repaired via HDR on the non-imprinted transcriptionally active allele either. Therefore, our primary hypothesis was rejected. Nevertheless, there are two possible explanations for our observations.

Firstly, HDR with a single-stranded repair should not be mistaken with traditional, Rad51-dependent HR. The mechanism of HDR with ssODNs requires Rad52, which is necessary for SSA in prokaryotes and yeast. In addition, it has also been proposed that HDR with ssODNs resembles SDSA and does not require incorporation of the donor molecule (Kan *et al.*, 2017). A link between HDR and the Fanconi Anemia pathway, which is responsible for the repair of inter-strand crosslinks, has also recently been established (Richardson *et al.*, 2018). This is not surprising, considering that Cas9 has been repeatedly reported to remain at DNA ends after having induced a DSB (Richardson *et al.*, 2016; Brinkman *et al.*, 2018). It is likely that the lack of distinct HDR efficiencies on the imprinted alleles is owed to the fundamentally

different DSB repair mechanism that was employed to repair Cas9-induced DSBs with the aid of an ssODN.

Secondly, the repair kinetics of Cas9-induced DSBs are reportedly slower than for IR-induced DSBs (Brinkman *et al.*, 2018). It is possible that the nature of Cas9-induced and IR-induced breaks is fundamentally different, especially since Cas9 may remain bound to DNA ends after cleavage or even remodel nucleosomes bearing any pre-existing histone modifications during target search by the Cas9-sgRNA complex. This possibility remains open to further investigation.

Due to reporting that chromatin compaction does not influence the choice between NHEJ and HDR during genome editing with CRISPR/Cas9, we have not managed to identify a chromatin state that is more permissive for repair of Cas9-induced DSBs via HDR. However, a number of different approaches have been undertaken to increase HDR frequency and ultimately result in higher rates of incorporation of the sequence information from the supplied repair templates. The most notable mentions include the use of plasmids to ectopically express RAD52 in HEK293 cells (Paulsen *et al.*, 2017) and an engineered version of RAD18, a RAD51 paralog, which reportedly increases HDR in HEK293 and HeLa cells (Nambiar *et al.*, 2019). Alternatively, Cas9 can be fused to short peptides that remodel chromatin (Ding *et al.*, 2019) or promote HDR, such as the N-terminal fragment of CtIP (Charpentier *et al.*, 2018). This research area will be particularly important in the future and will most likely shape genome editing with CRISPR/Cas9 into a better still method for targeted mutagenesis and introduction of desirable modifications to genomic locations of interest.

6.5 Concluding remarks

In summary, during my PhD I demonstrated that genomic imprinting can be successfully repurposed into an epigenetic system to assess how chromatin regulates CRISPR/Cas9 mutagenesis and repair of Cas9-induced DSBs with an externally provided single-stranded repair template. The key advantage of our approach is the simultaneous assessment of the same DNA sequence when it is compacted into two different chromatin states as defined by differential accessibility, DNA methylation and histone modifications. Our findings suggest that the efficiency of editing with CRISPR/Cas9 on a desirable target site can be increased if the site is located within accessible chromatin, a low concentration of Cas9 is used for a shortly-timed exposure, and the predicted off-target sites are located within inaccessible chromatin. In the future, we hope that our findings will be considered to refine tools for CRISPR design and pave the way towards the development of new off-target site prediction methods and improved strategies for genome editing for treating human disease.

References

Adachi, N., Nishijima, H. and Shibahara, K.-I. (2008) *Gene targeting using the human Nalm-6 pre-B cell line*, *BioScience Trends*.

Allen, F. *et al.* (2018) 'Predicting the mutations generated by repair of Cas9-induced double-strand breaks', *Nature Biotechnology*. Nature Publishing Group, 37(1), pp. 64–72. doi: 10.1038/nbt.4317.

Allen, N. D. *et al.* (1995) 'Distribution of parthenogenetic cells in the mouse brain and their influence on brain development and behavior.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 92(23), pp. 10782–6. doi: 10.1073/pnas.92.23.10782.

Allfrey, V. G., Faulkner, R. and Mirsky, A. E. (1964) 'Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis.', *Proceedings of the National Academy of Sciences*, 51(5), pp. 786–794.

Andres, S. N. *et al.* (2012) 'A human XRCC4–XLF complex bridges DNA', *Nucleic Acids Research*, 40(4), pp. 1868–1878. doi: 10.1093/nar/gks022.

Andresini, O. *et al.* (2019) 'The long non-coding RNA Kcnq1ot1 controls maternal p57 expression in muscle cells by promoting H3K27me3 accumulation to an intragenic MyoD-binding region', *Epigenetics & Chromatin*. BioMed Central, 12(1), p. 8. doi: 10.1186/s13072-019-0253-1.

Arakawa, H. and Iliakis, G. (2015) 'Alternative Okazaki Fragment Ligation Pathway by DNA Ligase III', *Genes*. Multidisciplinary Digital Publishing Institute, 6(2), pp. 385–398. doi: 10.3390/genes6020385.

Audebert, M., Salles, B. and Calsou, P. (2004) 'Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining.', *The Journal of biological chemistry*.

American Society for Biochemistry and Molecular Biology, 279(53), pp. 55117–26. doi: 10.1074/jbc.M404524200.

Aymard, F. *et al.* (2014) ‘Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks’, *Nature Structural & Molecular Biology*, 21(4), pp. 366–374. doi: 10.1038/nsmb.2796.

Baltimore, D. *et al.* (2015) ‘A prudent path forward for genomic engineering and germline gene modification’, *Science (New York, N.Y.)*. NIH Public Access, 348(6230), p. 36. doi: 10.1126/SCIENCE.AAB1028.

Bannister, A. J. *et al.* (2001) ‘Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain’, *Nature*. Nature Publishing Group, 410(6824), pp. 120–124. doi: 10.1038/35065138.

Bar, S. *et al.* (2017) ‘Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells’, *Cell Reports*. Elsevier B.V., 19(5), pp. 957–968. doi: 10.1016/j.celrep.2017.04.020.

Barkal, A. A. *et al.* (2016) ‘Cas9 functionally opens chromatin’, *PLoS ONE*, 11(3), pp. 1–8. doi: 10.1371/journal.pone.0152683.

Barlow, D. P. *et al.* (1991) ‘The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus’, *Nature*. Nature Publishing Group, 349(6304), pp. 84–87. doi: 10.1038/349084a0.

Barlow, D. P. (2011) ‘Genomic Imprinting: A Mammalian Epigenetic Discovery Model’, *Annual Review of Genetics*. Annual Reviews , 45(1), pp. 379–403. doi: 10.1146/annurev-genet-110410-132459.

Barlow, D. P. and Bartolomei, M. S. (2014) ‘Genomic imprinting in mammals.’, *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 6(2), p. a018382. doi: 10.1101/cshperspect.a018382.

Barrangou, R. *et al.* (2007) ‘CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes’, *Science*. American Association for the Advancement of Science, 315(5819), pp. 1709–1712. doi: 10.1126/science.1138140.

Bartolomei, M. S., Zemel, S. and Tilghman, S. M. (1991) ‘Parental imprinting of the mouse H19 gene’, *Nature*. Nature Publishing Group, 351(6322), pp. 153–155. doi: 10.1038/351153a0.

Baylin, S. B. and Schuebel, K. E. (2007) ‘The epigenomic era opens’, *Nature*. Nature Publishing Group, 448(7153), pp. 548–549. doi: 10.1038/448548a.

Bell, A. C. and Felsenfeld, G. (2000) ‘Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene’, *Nature*. Nature Publishing Group, 405(6785), pp. 482–485. doi: 10.1038/35013100.

Bell, A. C., West, A. G. and Felsenfeld, G. (1999) ‘The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.’, *Cell*. Elsevier, 98(3), pp. 387–96. doi: 10.1016/s0092-8674(00)81967-4.

Bernstein, B. E. *et al.* (2006) ‘A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells’, *Cell*. Cell Press, 125(2), pp. 315–326. doi: 10.1016/J.CELL.2006.02.041.

Bhaya, D., Davison, M. and Barrangou, R. (2011) ‘CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation’, *Annual Review of Genetics*. Annual Reviews , 45(1), pp. 273–297. doi: 10.1146/annurev-genet-110410-132430.

Bibikova, M. *et al.* (2001) ‘Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases’, *Molecular and Cellular Biology*. American Society for Microbiology Journals, 21(1), pp. 289–297. doi: 10.1128/MCB.21.1.289-297.2001.

Bibikova, M. *et al.* (2003) ‘Enhancing Gene Targeting with Designed Zinc Finger Nucleases’, *Science*, 300(5620), pp. 764–764. doi: 10.1126/science.1079512.

Birger, Y. *et al.* (1999) ‘The imprinting box of the mouse Igf2r gene’, *Nature*. Nature Publishing Group, 397(6714), pp. 84–86. doi: 10.1038/16291.

Bischof, O. *et al.* (2001) ‘Regulation and localization of the Bloom syndrome protein in response to DNA damage.’, *The Journal of cell biology*. Rockefeller

University Press, 153(2), pp. 367–80. doi: 10.1083/jcb.153.2.367.

Blaschke, K. *et al.* (2013) 'Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells', *Nature*. Nature Publishing Group, 500(7461), pp. 222–226. doi: 10.1038/nature12362.

Boch, J. and Bonas, U. (2010) '*Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function', *Annual Review of Phytopathology*. Annual Reviews , 48(1), pp. 419–436. doi: 10.1146/annurev-phyto-080508-081936.

Bogdanove, A. J., Schornack, S. and Lahaye, T. (2010) 'TAL effectors: finding plant genes for disease and defense', *Current Opinion in Plant Biology*. Elsevier Current Trends, 13(4), pp. 394–401. doi: 10.1016/J.PBI.2010.04.010.

Bonev, B. *et al.* (2017) 'Multiscale 3D Genome Rewiring during Mouse Neural Development', *Cell*, 171, pp. 557.e1-557.e24. doi: 10.1016/j.cell.2017.09.043.

Bostick, M. *et al.* (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*. American Association for the Advancement of Science, 317(5845), pp. 1760–1764. doi: 10.1126/science.1147939.

Bothmer, A. *et al.* (2017) 'Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus', *Nature Communications*, 8. doi: 10.1038/ncomms13905.

Bouabe, H. and Okkenhaug, K. (2013) 'Gene Targeting in Mice: A Review', in. Humana Press, Totowa, NJ, pp. 315–336. doi: 10.1007/978-1-62703-601-6_23.

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the establishment of maternal genomic imprints.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 294(5551), pp. 2536–9. doi: 10.1126/science.1065848.

Bouschet, T. *et al.* (2017) 'In Vitro Corticogenesis from Embryonic Stem Cells

Recapitulates the In Vivo Epigenetic Control of Imprinted Gene Expression', *Cerebral cortex (New York, N.Y. : 1991)*, 27(3), pp. 2418–2433. doi: 10.1093/cercor/bhw102.

Brenner, S. (1974) 'The genetics of Caenorhabditis elegans', *Genetics*, 77(1), pp. 71–94.

Brinkman, E. K. *et al.* (2018) 'Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks', *Molecular Cell*, 70(5). doi: 10.1016/j.molcel.2018.04.016.

Brown, S. W. (1966) 'Heterochromatin', *Science*, 151(3709), pp. 417–425. Available at: https://www.jstor.org/stable/1717239?seq=2#metadata_info_tab_contents (Accessed: 15 August 2019).

Bugreev, D. V, Hanaoka, F. and Mazin, A. V (2007) 'Rad54 dissociates homologous recombination intermediates by branch migration', *Nature Structural & Molecular Biology*. Nature Publishing Group, 14(8), pp. 746–753. doi: 10.1038/nsmb1268.

Burman, B. *et al.* (2015) 'Histone modifications predispose genome regions to breakage and translocation.', *Genes & development*. Cold Spring Harbor Laboratory Press, 29(13), pp. 1393–402. doi: 10.1101/gad.262170.115.

Cameron, P. *et al.* (2017) 'SITE-Seq: A Genome-wide Method to Measure Cas9 Cleavage', *Protocol Exchange*. doi: 10.1038/protex.2017.043.

Cao, R. *et al.* (2002) 'Role of histone H3 lysine 27 methylation in Polycomb-group silencing.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 298(5595), pp. 1039–43. doi: 10.1126/science.1076997.

Capecchi, M. R. (2001) 'Generating mice with targeted mutations', *Nature Medicine*. Nature Publishing Group, 7(10), pp. 1086–1090. doi: 10.1038/nm1001-1086.

Capecchi, M. R. (2005) 'Gene targeting in mice: functional analysis of the

mammalian genome for the twenty-first century', *Nature Reviews Genetics*. Nature Publishing Group, 6(6), pp. 507–512. doi: 10.1038/nrg1619.

Cattanach, B. M. (1986) 'Parental origin effects in mice', *Development*, 97, pp. 137–150.

Cattanach, B. M. and Kirk, M. (1985) 'Differential activity of maternally and paternally derived chromosome regions in mice', *Nature*. Nature Publishing Group, 315(6019), pp. 496–498. doi: 10.1038/315496a0.

Cebrian-Serrano, A. and Davies, B. (2017) 'CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools', *Mammalian Genome*. Springer US, 28(7–8), pp. 247–261. doi: 10.1007/s00335-017-9697-4.

Ceccaldi, R., Rondinelli, B. and D'Andrea, A. D. (2016) 'Repair Pathway Choices and Consequences at the Double-Strand Break', *Trends in Cell Biology*. Elsevier Current Trends, 26(1), pp. 52–64. doi: 10.1016/J.TCB.2015.07.009.

Cermak, T. *et al.* (2011) 'Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting', *Nucleic Acids Research*, 39(12), pp. e82–e82. doi: 10.1093/nar/gkr218.

Chakrabarti, A. M. *et al.* (2018) 'Target-Specific Precision of CRISPR-Mediated Genome Editing.', *Molecular cell*. Elsevier, 0(0). doi: 10.1016/j.molcel.2018.11.031.

Chan, D. W. *et al.* (2002) 'Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks.', *Genes & development*. Cold Spring Harbor Laboratory Press, 16(18), pp. 2333–8. doi: 10.1101/gad.1015202.

Char, S. N. *et al.* (2017) 'An *Agrobacterium* -delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize', *Plant Biotechnology Journal*. John Wiley & Sons, Ltd (10.1111), 15(2), pp. 257–268. doi: 10.1111/pbi.12611.

Chari, R. *et al.* (2015) ‘Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach’, *Nature Methods*. Nature Publishing Group, 12(9), pp. 823–826. doi: 10.1038/nmeth.3473.

Charpentier, M. *et al.* (2018) ‘CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair’, *Nature Communications*. Nature Publishing Group, 9(1), p. 1133. doi: 10.1038/s41467-018-03475-7.

Chen, S. *et al.* (2015) ‘Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis’, *Cell*. Elsevier, 160(6), pp. 1246–1260. doi: 10.1016/J.CELL.2015.02.038.

Chen, W. *et al.* (2019) ‘Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair’, *Nucleic Acids Research*. doi: 10.1093/nar/gkz487.

Chen, X. *et al.* (2016) ‘Probing the impact of chromatin conformation on genome editing tools’, *Nucleic Acids Research*, 44(13), pp. 6482–6492. doi: 10.1093/nar/gkw524.

Chen, Y. *et al.* (2017) ‘Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish’, *PLoS ONE*, 12(8). doi: 10.1371/journal.pone.0182528.

Chiolo, I. *et al.* (2011) ‘Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair’, *Cell*. Elsevier Inc., 144(5), pp. 732–744. doi: 10.1016/j.cell.2011.02.012.

Cho, S. W. *et al.* (2013) ‘Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease’, *Nature Biotechnology*. Nature Publishing Group, 31(3), pp. 230–232. doi: 10.1038/nbt.2507.

Choi, J. *et al.* (2017) ‘Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells’, *Nature*. Nature Publishing Group, 548(7666), pp. 219–223. doi: 10.1038/nature23274.

Choi, J. D. *et al.* (2005) ‘A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG

island.', *Molecular and cellular biology*. American Society for Microbiology Journals, 25(13), pp. 5514–22. doi: 10.1128/MCB.25.13.5514-5522.2005.

Choi, P. S. and Meyerson, M. (2014) 'Targeted genomic rearrangements using CRISPR/Cas technology', *Nature Communications*. Nature Publishing Group, 5(1), p. 3728. doi: 10.1038/ncomms4728.

Choulika, A. *et al.* (1995) 'Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 15(4), pp. 1968–73. doi: 10.1128/mcb.15.4.1968.

Chung, J. E. *et al.* (2019) 'CRISPR-Cas9 interrogation of a putative fetal globin repressor in human erythroid cells', *PLOS ONE*. Edited by A. C. Wilber. Public Library of Science, 14(1), p. e0208237. doi: 10.1371/journal.pone.0208237.

Ciccone, D. N. *et al.* (2009) 'KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints', *Nature*. Nature Publishing Group, 461(7262), pp. 415–418. doi: 10.1038/nature08315.

Clouaire, T. and Legube, G. (2015) 'DNA double strand break repair pathway choice: a chromatin based decision?', *Nucleus*, 6(2), pp. 107–113. doi: 10.1080/19491034.2015.1010946.

Coan, P. M. *et al.* (2008) 'Disproportional effects of Igf2 knockout on placental morphology and diffusional exchange characteristics in the mouse', *Journal of Physiology*. Wiley-Blackwell, 586(20), pp. 5023–5032. doi: 10.1113/jphysiol.2008.157313.

Coan, P. M., Burton, G. J. and Ferguson-Smith, A. C. (2005) 'Imprinted genes in the placenta – A review', *Placenta*. W.B. Saunders, 26, pp. S10–S20. doi: 10.1016/J.PLACENTA.2004.12.009.

Comings, D. E. (1978) 'Mechanisms of Chromosome Banding and Implications for Chromosome Structure', *Annual Review of Genetics*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139,

USA , 12(1), pp. 25–46. doi: 10.1146/annurev.ge.12.120178.000325.

Cong, L. *et al.* (2013) ‘Multiplex genome engineering using CRISPR/Cas systems.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 339(6121), pp. 819–23. doi: 10.1126/science.1231143.

Craig, J. M. and Bickmore, W. A. (1993) ‘Genes and genomes: Chromosome bands - flavours to savour’, *BioEssays*. John Wiley & Sons, Ltd, 15(5), pp. 349–354. doi: 10.1002/bies.950150510.

Cubbon, A., Ivancic-Bace, I. and Bolt, E. L. (2018) ‘CRISPR-Cas immunity, DNA repair and genome stability.’, *Bioscience reports*. Portland Press Limited, 38(5), p. BSR20180457. doi: 10.1042/BSR20180457.

Daer, R. M. *et al.* (2017) ‘The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells’, *ACS Synthetic Biology*. American Chemical Society, 6(3), pp. 428–438. doi: 10.1021/acssynbio.5b00299.

Dagdas, Y. S. *et al.* (2017) ‘A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9’, *Science Advances*. American Association for the Advancement of Science, 3(8), p. eaao0027. doi: 10.1126/sciadv.aao0027.

Daley, J. M. *et al.* (2005) ‘DNA joint dependence of pol X family polymerase action in nonhomologous end joining.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(32), pp. 29030–7. doi: 10.1074/jbc.M505277200.

Daugaard, M. *et al.* (2012) ‘LEDGF (p75) promotes DNA-end resection and homologous recombination’, *Nature Structural and Molecular Biology*. Nature Publishing Group, 19(8), pp. 803–810. doi: 10.1038/nsmb.2314.

Davis, T. L. *et al.* (2000) ‘The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development’, *Human Molecular Genetics*, 9(19), pp. 2885–2894. doi: 10.1093/hmg/9.19.2885.

DeChiara, T. M., Efstratiadis, A. and Robertsen, E. J. (1990) ‘A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting’, *Nature*. Nature Publishing Group, 345(6270), pp. 78–80. doi: 10.1038/345078a0.

DeChiara, T. M., Robertson, E. J. and Efstratiadis, A. (1991) ‘Parental imprinting of the mouse insulin-like growth factor II gene’, *Cell*. Cell Press, 64(4), pp. 849–859. doi: 10.1016/0092-8674(91)90513-X.

Dennis, K. *et al.* (2001) ‘Lsh, a member of the SNF2 family, is required for genome-wide methylation’, *Genes and Development*. Cold Spring Harbor Laboratory Press, 15(22), pp. 2940–2944. doi: 10.1101/gad.929101.

DeWitt, M. A. *et al.* (2016) ‘Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells.’, *Science translational medicine*. American Association for the Advancement of Science, 8(360), p. 360ra134. doi: 10.1126/scitranslmed.aaf9336.

Dillon, N. and Festenstein, R. (2002) ‘Unravelling heterochromatin: competition between positive and negative factors regulates accessibility’, *Trends in Genetics*. Elsevier Current Trends, 18(5), pp. 252–258. doi: 10.1016/S0168-9525(02)02648-3.

Ding, X. *et al.* (2019) ‘Improving CRISPR-Cas9 Genome Editing Efficiency by Fusion with Chromatin-Modulating Peptides’, *The CRISPR Journal*. Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 2(1), pp. 51–63. doi: 10.1089/crispr.2018.0036.

Doench, J. G. *et al.* (2014) ‘Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation’, *Nature Biotechnology*. Nature Publishing Group, 32(12), pp. 1262–1267. doi: 10.1038/nbt.3026.

Doench, J. G. *et al.* (2016) ‘Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9’, *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 184–191. doi: 10.1038/nbt.3437.

Doetschman, T. *et al.* (1987) ‘Targetted correction of a mutant HPRT gene in

mouse embryonic stem cells', *Nature*. Nature Publishing Group, 330(6148), pp. 576–578. doi: 10.1038/330576a0.

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*. American Association for the Advancement of Science, 346(6213), pp. 1258096–1258096. doi: 10.1126/science.1258096.

Downs, J. A. and Jackson, S. P. (2004) 'A means to a DNA end: the many roles of Ku', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 5(5), pp. 367–378. doi: 10.1038/nrm1367.

Durai, S. *et al.* (2005) 'Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells', *Nucleic Acids Research*, 33(18), pp. 5978–5990. doi: 10.1093/nar/gki912.

Edwards, C. A. and Ferguson-Smith, A. C. (2007) 'Mechanisms regulating imprinted genes in clusters', *Current Opinion in Cell Biology*. Elsevier Current Trends, 19(3), pp. 281–289. doi: 10.1016/J.CEB.2007.04.013.

Eid, W. *et al.* (2010) 'DNA end resection by CtIP and exonuclease 1 prevents genomic instability', *EMBO reports*. John Wiley & Sons, Ltd, 11(12), pp. 962–968. doi: 10.1038/embor.2010.157.

Engel, N. *et al.* (2004) 'Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations', *Nature Genetics*. Nature Publishing Group, 36(8), pp. 883–888. doi: 10.1038/ng1399.

Epsztejn-Litman, S. *et al.* (2008) 'De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes', *Nature Structural and Molecular Biology*. Nature Publishing Group, 15(11), pp. 1176–1183. doi: 10.1038/nsmb.1476.

Esashi, F. *et al.* (2005) 'CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair', *Nature*. Nature Publishing Group, 434(7033), pp. 598–604. doi: 10.1038/nature03404.

Estandarte, A. K. *et al.* (2016) ‘The use of DAPI fluorescence lifetime imaging for investigating chromatin condensation in human chromosomes.’, *Scientific reports*. Nature Publishing Group, 6, p. 31417. doi: 10.1038/srep31417.

Estève, P. O. *et al.* (2006) ‘Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication’, *Genes and Development*. Cold Spring Harbor Laboratory Press, 20(22), pp. 3089–3103. doi: 10.1101/gad.1463706.

Fawcett, D. W. (1966) ‘On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates’, *American Journal of Anatomy*, 119(1), pp. 129–145. doi: 10.1002/aja.1001190108.

Feil, R. *et al.* (1997) ‘Parental chromosome-specific chromatin conformation in the imprinted U2af1-rs1 gene in the mouse.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 272(33), pp. 20893–900. doi: 10.1074/jbc.272.33.20893.

De Felipe, P. *et al.* (2006) ‘E unum pluribus: Multiple proteins from a self-processing polyprotein’, *Trends in Biotechnology*. Elsevier Current Trends, pp. 68–75. doi: 10.1016/j.tibtech.2005.12.006.

Feng, H. P., Scherl, D. S. and Widom, J. (1993) ‘Lifetime of the histone octamer studied by continuous-flow quasielastic light scattering: Test of a model for nucleosome transcription’, *Biochemistry*, 32(30), pp. 7824–7831. doi: 10.1021/bi00081a030.

Ferguson-Smith, A. C. *et al.* (1991) ‘Embryological and molecular investigations of parental imprinting on mouse chromosome 7’, *Nature*. Nature Publishing Group, 351(6328), pp. 667–670. doi: 10.1038/351667a0.

Ferguson-Smith, A. C. (2011) ‘Genomic imprinting: the emergence of an epigenetic paradigm’, *Nature Reviews Genetics*. Nature Publishing Group, 12(8), pp. 565–575. doi: 10.1038/nrg3032.

Ferguson-Smith, A. C. and Bourchis, D. (2018) ‘The discovery and importance of genomic imprinting’, *eLife*, 7. doi: 10.7554/eLife.42368.

Filippova, G. N. *et al.* (1996) ‘An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes.’, *Molecular and cellular biology*. American Society for Microbiology Journals, 16(6), pp. 2802–13. doi: 10.1128/mcb.16.6.2802.

Filson, A. J. *et al.* (1993) ‘Rescue of the T-associated maternal effect in mice carrying null mutations in Igf-2 and Igf2r, two reciprocally imprinted genes.’, *Development* . The Company of Biologists Ltd, 118(3), pp. 731–6.

Fischle, W. *et al.* (2005) ‘Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation’, *Nature*. Nature Publishing Group, 438(7071), pp. 1116–1122. doi: 10.1038/nature04219.

Fishman-Lobell, J., Rudin, N. and Haber, J. E. (1992) ‘Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated.’, *Molecular and Cellular Biology*. American Society for Microbiology Journals, 12(3), pp. 1292–1303. doi: 10.1128/MCB.12.3.1292.

Fitzpatrick, G. V., Soloway, P. D. and Higgins, M. J. (2002) ‘Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1’, *Nature Genetics*. Nature Publishing Group, 32(3), pp. 426–431. doi: 10.1038/ng988.

Fournier, C. *et al.* (2002) ‘Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes’, *The EMBO Journal*. John Wiley & Sons, Ltd, 21(23), pp. 6560–6570. doi: 10.1093/emboj/cdf655.

Frommer, M. *et al.* (1992) ‘A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.’, *Genetics*, 89, pp. 1827–1831.

Fu, F. and Voytas, D. F. (2012) ‘Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C2H2 zinc fingers and engineered zinc finger arrays’, *Nucleic Acids Research*, 41(D1), pp. D452–D455. doi: 10.1093/nar/gks1167.

Fujita, T., Yuno, M. and Fujii, H. (2016) ‘Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus’, *Scientific Reports*. Nature Publishing Group, 6(1), p. 30485. doi: 10.1038/srep30485.

Fuks, F. *et al.* (2000) ‘DNA methyltransferase Dnmt1 associates with histone deacetylase activity’, *Nature Genetics*. Nature Publishing Group, 24(1), pp. 88–91. doi: 10.1038/71750.

Gallagher, D. N. and Haber, J. E. (2018) ‘Repair of a Site-Specific DNA Cleavage: Old-School Lessons for Cas9-Mediated Gene Editing’, *ACS Chemical Biology*. American Chemical Society, 13(2), pp. 397–405. doi: 10.1021/acschembio.7b00760.

Garfield, A. S. *et al.* (2011) ‘Distinct physiological and behavioural functions for parental alleles of imprinted Grb10’, *Nature*. Nature Publishing Group, 469(7331), pp. 534–538. doi: 10.1038/nature09651.

Garg, A. *et al.* (2012) ‘Engineering synthetic TAL effectors with orthogonal target sites’, *Nucleic Acids Research*, 40(15), pp. 7584–7595. doi: 10.1093/nar/gks404.

Garneau, J. E. *et al.* (2010) ‘The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA’, *Nature*. Nature Publishing Group, 468(7320), pp. 67–71. doi: 10.1038/nature09523.

George, S. H. L. *et al.* (2007) ‘Developmental and adult phenotyping directly from mutant embryonic stem cells’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(11), pp. 4455–4460. doi: 10.1073/pnas.0609277104.

Giannoukakis, N. *et al.* (1993) ‘Parental genomic imprinting of the human IGF2 gene’, *Nature Genetics*. Nature Publishing Group, 4(1), pp. 98–101. doi: 10.1038/ng0593-98.

Glaser, S., Anastassiadis, K. and Stewart, A. F. (2005) ‘Current issues in mouse genome engineering’, *Nature Genetics*. Nature Publishing Group, 37(11), pp. 1187–1193. doi: 10.1038/ng1668.

Globyte, V. *et al.* (2019) ‘CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion’, *The EMBO Journal*. EMBO Press, 38(4), p. e99466. doi: 10.15252/embj.201899466.

Gong, S. *et al.* (2018) ‘DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity.’, *Cell reports*. Elsevier, 22(2), pp. 359–371. doi: 10.1016/j.celrep.2017.12.041.

Goodarzi, A. A. *et al.* (2006) ‘DNA-PK autophosphorylation facilitates Artemis endonuclease activity’, *The EMBO Journal*. John Wiley & Sons, Ltd, 25(16), pp. 3880–3889. doi: 10.1038/sj.emboj.7601255.

Greenberg, M. V. C. and Bourc’his, D. (2019) ‘The diverse roles of DNA methylation in mammalian development and disease’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 590–607. doi: 10.1038/s41580-019-0159-6.

Gupta, A. *et al.* (2011) ‘Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases’, *Nucleic Acids Research*, 39(1), pp. 381–392. doi: 10.1093/nar/gkq787.

Haince, J.-F. *et al.* (2008) ‘PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 283(2), pp. 1197–208. doi: 10.1074/jbc.M706734200.

Hark, A. T. *et al.* (2000) ‘CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus’, *Nature*. Nature Publishing Group, 405(6785), pp. 486–489. doi: 10.1038/35013106.

He, Y. F. *et al.* (2011) ‘Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA’, *Science*. American Association for the Advancement of Science, 333(6047), pp. 1303–1307. doi: 10.1126/science.1210944.

Heard, E. and Disteche, C. M. (2006) ‘Dosage compensation in mammals: fine-tuning the expression of the X chromosome.’, *Genes & development*.

Cold Spring Harbor Laboratory Press, 20(14), pp. 1848–67. doi: 10.1101/gad.1422906.

Hebbes, T. R., Thorne, A. W. and Crane-Robinson, C. (1988) ‘A direct link between core histone acetylation and transcriptionally active chromatin.’, *The EMBO Journal*. John Wiley & Sons, Ltd, 7(5), pp. 1395–1402. doi: 10.1002/j.1460-2075.1988.tb02956.x.

Héberlé, É. and Bardet, A. F. (2019) ‘Sensitivity of transcription factors to DNA methylation’, *Essays in Biochemistry*. Portland Press Ltd, pp. 727–741. doi: 10.1042/EBC20190033.

Henckel, A. *et al.* (2009) ‘Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals’, *Human Molecular Genetics*, 18(18), pp. 3375–3383. doi: 10.1093/hmg/ddp277.

Hendrich, B. and Bird, A. (1998) ‘Identification and characterization of a family of mammalian methyl CpG-binding proteins’, *Genetical Research*. Cambridge University Press (CUP), 72(1), pp. 59–72. doi: 10.1017/s0016672398533307.

Hermann, M. *et al.* (2014) ‘Mouse Genome Engineering Using Designer Nucleases’, *Journal of Visualized Experiments*, (86), p. e50930. doi: 10.3791/50930.

Hinz, J. M., Laughery, M. F. and Wyrick, J. J. (2015) ‘Nucleosomes Inhibit Cas9 Endonuclease Activity *in Vitro*’, *Biochemistry*. American Chemical Society, 54(48), pp. 7063–7066. doi: 10.1021/acs.biochem.5b01108.

Hinz, J. M., Laughery, M. F. and Wyrick, J. J. (2016) ‘Nucleosomes Selectively Inhibit Cas9 Off-target Activity at a Site Located at the Nucleosome Edge.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 291(48), pp. 24851–24856. doi: 10.1074/jbc.C116.758706.

Hockemeyer, D. *et al.* (2009) ‘Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases’, *Nature*

Biotechnology. Nature Publishing Group, 27(9), pp. 851–857. doi: 10.1038/nbt.1562.

Hockemeyer, D. *et al.* (2011) 'Genetic engineering of human pluripotent cells using TALE nucleases', *Nature Biotechnology*. Nature Publishing Group, 29(8), pp. 731–734. doi: 10.1038/nbt.1927.

Horlbeck, M. A. *et al.* (2016) 'Nucleosomes impede Cas9 access to DNA in vivo and in vitro', *eLife*, 5. doi: 10.7554/eLife.12677.

Horvath, P. and Barrangou, R. (2010) 'CRISPR/Cas, the immune system of bacteria and archaea.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 327(5962), pp. 167–70. doi: 10.1126/science.1179555.

Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and Applications of CRISPR-Cas9 for Genome Engineering', *Cell*. Cell Press, 157(6), pp. 1262–1278. doi: 10.1016/J.CELL.2014.05.010.

Iliakis, G. *et al.* (2004) 'Mechanisms of DNA double strand break repair and chromosome aberration formation', *Cytogenetic and Genome Research*. Karger Publishers, 104(1–4), pp. 14–20. doi: 10.1159/000077461.

Inagaki, A., Schoenmakers, S. and Baarends, W. M. (2010) 'DNA double strand break repair, chromosome synapsis and transcriptional silencing in meiosis', *Epigenetics*. Taylor & Francis, 5(4), pp. 255–266. doi: 10.4161/epi.5.4.11518.

Ip, S. C. Y. *et al.* (2008) 'Identification of Holliday junction resolvases from humans and yeast', *Nature*. Nature Publishing Group, 456(7220), pp. 357–361. doi: 10.1038/nature07470.

Isaac, R. S. *et al.* (2016) 'Nucleosome breathing and remodeling constrain CRISPR-Cas9 function', *eLife*, 5(APRIL2016), pp. 1–14. doi:

10.7554/eLife.13450.

Ishiyama, S. *et al.* (2017) ‘Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance’, *Molecular Cell*. Cell Press, 68(2), pp. 350-360.e7. doi: 10.1016/j.molcel.2017.09.037.

Ito, S. *et al.* (2010) ‘Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification’, *Nature*. Nature Publishing Group, 466(7310), pp. 1129–1133. doi: 10.1038/nature09303.

Ito, S. *et al.* (2011) ‘Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine’, *Science*. American Association for the Advancement of Science, 333(6047), pp. 1300–1303. doi: 10.1126/science.1210597.

Jakob, B. *et al.* (2011) ‘DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin’, *Nucleic Acids Research*, 39(15), pp. 6489–6499. doi: 10.1093/nar/gkr230.

Janssen, A. *et al.* (2016) ‘A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 30(14), pp. 1645–57. doi: 10.1101/gad.283028.116.

Janssen, J. M. *et al.* (2019) ‘The chromatin structure of CRISPR-Cas9 target sequences controls the balance between mutagenic and homology-directed gene editing events’, *Molecular Therapy - Nucleic Acids*. Elsevier, 0(0). doi: 10.1016/j.omtn.2019.02.009.

Jasin, M. (1996) ‘Genetic manipulation of genomes with rare-cutting endonucleases’, *Trends in Genetics*. Elsevier Current Trends, 12(6), pp. 224–228. doi: 10.1016/0168-9525(96)10019-6.

Jensen, K. T. *et al.* (2017) ‘Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency’, *FEBS*

Letters. John Wiley & Sons, Ltd, 591(13), pp. 1892–1901. doi: 10.1002/1873-3468.12707.

Jiang, F. *et al.* (2015) 'STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 348(6242), pp. 1477–81. doi: 10.1126/science.aab1452.

Jiang, F. *et al.* (2016) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 351(6275), pp. 867–71. doi: 10.1126/science.aad8282.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*, 2. doi: 10.7554/eLife.00471.

Jinek, M. *et al.* (2014) 'Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 343(6176), p. 1247997. doi: 10.1126/science.1247997.

Joung, J. K. and Sander, J. D. (2013) 'TALENs: a widely applicable technology for targeted genome editing', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 14(1), pp. 49–55. doi: 10.1038/nrm3486.

Kallimasioti-Pazi, E. M. *et al.* (2018) 'Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair', *PLoS Biology*. doi: 10.1101/267690.

Kan, Y. *et al.* (2017) 'Mechanisms of precise genome editing using oligonucleotide donors.', *Genome research*. Cold Spring Harbor Laboratory Press, 27(7), pp. 1099–1111. doi: 10.1101/gr.214775.116.

Kaneda, M. *et al.* (2004) 'Essential role for de novo DNA methyltransferase

Dnmt3a in paternal and maternal imprinting', *Nature*. Nature Publishing Group, 429(6994), pp. 900–903. doi: 10.1038/nature02633.

Kass, E. M. *et al.* (2016) 'Robust homology-directed repair within mouse mammary tissue is not specifically affected by Brca2 mutation', *Nature Communications*. Nature Publishing Group, 7(1), p. 13241. doi: 10.1038/ncomms13241.

Kim, D. *et al.* (2016) 'Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq.', *Genome research*. Cold Spring Harbor Laboratory Press, 26(3), pp. 406–15. doi: 10.1101/gr.199588.115.

Kim, D. and Kim, J.-S. (2018) 'DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA.', *Genome research*. Cold Spring Harbor Laboratory Press, 28(12), pp. 1894–1900. doi: 10.1101/gr.236620.118.

Kim, S. *et al.* (2014) 'Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins', *Genome Research*, 24(6), pp. 1012–1019. doi: 10.1101/gr.171322.113.

Kim, Y. G., Cha, J. and Chandrasegaran, S. (1996) 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 93(3), pp. 1156–60. doi: 10.1073/pnas.93.3.1156.

Knezetic, J. A. and Luse, D. S. (1986) 'The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro.', *Cell*. Elsevier, 45(1), pp. 95–104. doi: 10.1016/0092-8674(86)90541-6.

Knight, S. C. *et al.* (2015) 'Dynamics of CRISPR-Cas9 genome interrogation in living cells.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 350(6262), pp. 823–6. doi: 10.1126/science.aac6572.

Koike-Yusa, H. *et al.* (2014) 'Genome-wide recessive genetic screening in

mammalian cells with a lentiviral CRISPR-guide RNA library', *Nature Biotechnology*. Nature Publishing Group, 32(3), pp. 267–273. doi: 10.1038/nbt.2800.

Komor, A. C., Badran, A. H. and Liu, D. R. (2017) 'CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes', *Cell*. Cell Press, 168(1–2), pp. 20–36. doi: 10.1016/J.CELL.2016.10.044.

Kornberg, R. D. (1974) 'Chromatin structure: a repeating unit of histones and DNA.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 184(4139), pp. 868–71. doi: 10.1126/science.184.4139.868.

Kornberg, R. D. and Thonmas, J. O. (1974) 'Chromatin structure: a repeating unit of histones and DNA.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 184(4139), pp. 868–71. doi: 10.1126/science.184.4139.868.

Kosicki, M., Tomberg, K. and Bradley, A. (2018) 'Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements', *Nature Biotechnology*. Nature Publishing Group, 36(8), p. 765. doi: 10.1038/nbt.4192.

Kota, S. K. *et al.* (2014) 'ICR Noncoding RNA Expression Controls Imprinting and DNA Replication at the Dlk1-Dio3 Domain', *Developmental Cell*. Cell Press, 31(1), pp. 19–33. doi: 10.1016/J.DEVCEL.2014.08.009.

Kraft, K. *et al.* (2015) 'Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice', *Cell Reports*. Cell Press, 10(5), pp. 833–839. doi: 10.1016/J.CELREP.2015.01.016.

Krogan, N. J. *et al.* (2003) 'Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.', *Molecular and cellular biology*. American Society for Microbiology Journals, 23(12), pp. 4207–18. doi: 10.1128/mcb.23.12.4207-4218.2003.

Krueger, F. and Andrews, S. R. (2011) 'Bismark: a flexible aligner and

methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27(11), pp. 1571–1572. doi: 10.1093/bioinformatics/btr167.

Kuscu, C. *et al.* (2014) 'Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease', *Nature Biotechnology*. Nature Publishing Group, 32(7), pp. 677–683. doi: 10.1038/nbt.2916.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*. Nature Publishing Group, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

LaRocque, J. R. and Jasin, M. (2010) 'Mechanisms of Recombination between Diverged Sequences in Wild-Type and BLM-Deficient Mouse and Human Cells', *Molecular and Cellular Biology*. American Society for Microbiology Journals, 30(8), pp. 1887–1897. doi: 10.1128/MCB.01553-09.

Larson, M. H. *et al.* (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nature Protocols*. Nature Publishing Group, 8(11), pp. 2180–2196. doi: 10.1038/nprot.2013.132.

Leitch, H. G. *et al.* (2013) 'Naive pluripotency is associated with global DNA hypomethylation', *Nature Structural and Molecular Biology*. Nature Publishing Group, 20(3), pp. 311–316. doi: 10.1038/nsmb.2510.

Lemaître, C. *et al.* (2014) 'Nuclear position dictates DNA repair pathway choice', *Genes & Development*. Cold Spring Harbor Laboratory Press, 28(22), pp. 2450–2463. doi: 10.1101/GAD.248369.114.

Li, B., Carey, M. and Workman, J. L. (2007) 'The Role of Chromatin during Transcription', *Cell*. Cell Press, 128(4), pp. 707–719. doi: 10.1016/J.CELL.2007.01.015.

Li, E., Beard, C. and Jaenisch, R. (1993) 'Role for DNA methylation in genomic imprinting', *Nature*. Nature Publishing Group, 366(6453), pp. 362–365. doi: 10.1038/366362a0.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows–Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi:

10.1093/bioinformatics/btp698.

Li, K. *et al.* (2014) ‘Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System’, *PLoS ONE*. Edited by K. Yu. Public Library of Science, 9(8), p. e105779. doi: 10.1371/journal.pone.0105779.

Li, T., Huang, S., Zhao, X., *et al.* (2011) ‘Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes’, *Nucleic Acids Research*, 39(14), pp. 6315–6325. doi: 10.1093/nar/gkr188.

Li, T., Huang, S., Jiang, W. Z., *et al.* (2011) ‘TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain.’, *Nucleic acids research*. Oxford University Press, 39(1), pp. 359–72. doi: 10.1093/nar/gkq704.

Li, X. *et al.* (2008) ‘A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints’, *Developmental Cell*. Cell Press, 15(4), pp. 547–557. doi: 10.1016/J.DEVCEL.2008.08.014.

Lieber, M. R. *et al.* (2003) ‘Mechanism and regulation of human non-homologous DNA end-joining’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 4(9), pp. 712–720. doi: 10.1038/nrm1202.

Lin, F. L., Sperle, K. and Sternberg, N. (1985) ‘Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 82(5), pp. 1391–5. doi: 10.1073/pnas.82.5.1391.

Lomova, A. *et al.* (2019) ‘Improving Gene Editing Outcomes in Human Hematopoietic Stem and Progenitor Cells by Temporal Control of DNA Repair’, *STEM CELLS*. John Wiley & Sons, Ltd, 37(2), pp. 284–294. doi: 10.1002/stem.2935.

Lorch, Y., LaPointe, J. W. and Kornberg, R. D. (1987) ‘Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement

of histones.', *Cell*. Elsevier, 49(2), pp. 203–10. doi: 10.1016/0092-8674(87)90561-7.

Lowary, P. . and Widom, J. (1998) 'New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning', *Journal of Molecular Biology*. Academic Press, 276(1), pp. 19–42. doi: 10.1006/JMBI.1997.1494.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution', *Nature*. Nature Publishing Group, 389(6648), pp. 251–260. doi: 10.1038/38444.

Mahfouz, M. M. *et al.* (2011) 'De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(6), pp. 2623–8. doi: 10.1073/pnas.1019533108.

Makarova, K. S. *et al.* (2011) 'Evolution and classification of the CRISPR–Cas systems', *Nature Reviews Microbiology*. Nature Publishing Group, 9(6), pp. 467–477. doi: 10.1038/nrmicro2577.

Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR–Cas systems', *Nature Reviews Microbiology*. Nature Publishing Group, 13(11), pp. 722–736. doi: 10.1038/nrmicro3569.

Makharashvili, N. and Paull, T. T. (2015) 'CtIP: A DNA damage response protein at the intersection of DNA metabolism', *DNA Repair*. Elsevier, 32, pp. 75–81. doi: 10.1016/J.DNAREP.2015.04.016.

Mali, P. *et al.* (2013) 'RNA-Guided Human Genome Engineering via Cas9', *Science*. American Association for the Advancement of Science, 339(6121), pp. 823–826. doi: 10.1126/SCIENCE.1232033.

Malkova, A. *et al.* (2001) 'RAD51-independent break-induced replication to repair a broken chromosome depends on a distant enhancer site.', *Genes & development*. Cold Spring Harbor Laboratory Press, 15(9), pp. 1055–60. doi:

10.1101/gad.875901.

Mancini-Dinardo, D. *et al.* (2006) ‘Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 20(10), pp. 1268–82. doi: 10.1101/gad.1416906.

Mann, J. R. *et al.* (1990) ‘Androgenetic mouse embryonic stem cells are pluripotent and cause skeletal defects in chimeras: Implications for genetic imprinting’, *Cell*. Cell Press, 62(2), pp. 251–260. doi: 10.1016/0092-8674(90)90363-J.

Mann, J. R. and Lovell-Badge, R. H. (1984) ‘Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm’, *Nature*. Nature Publishing Group, 310(5972), pp. 66–67. doi: 10.1038/310066a0.

Martens, J. H. A. *et al.* (2005) ‘The profile of repeat-associated histone lysine methylation states in the mouse epigenome’, *The EMBO Journal*. John Wiley & Sons, Ltd, 24(4), pp. 800–812. doi: 10.1038/sj.emboj.7600545.

Masson, J. Y. *et al.* (2001) ‘Identification and purification of two distinct complexes containing the five RAD51 paralogs.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 15(24), pp. 3296–307. doi: 10.1101/gad.947001.

Maupetit-Méhouas, S. *et al.* (2016) ‘Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive’, *Nucleic Acids Research*, 44(2), pp. 621–635. doi: 10.1093/nar/gkv960.

Mazin, A. V. *et al.* (2010) ‘Rad54, the motor of homologous recombination’, *DNA Repair*. Elsevier, 9(3), pp. 286–302. doi: 10.1016/J.DNAREP.2009.12.006.

Mazina, O. M. and Mazin, A. V (2008) ‘Human Rad54 protein stimulates human Mus81-Eme1 endonuclease.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(47), pp. 18249–54. doi: 10.1073/pnas.0807016105.

McArthur, M., Gerum, S. and Stamatoyannopoulos, G. (2001) ‘Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse β-globin LCR’, *Journal of Molecular Biology*. Academic Press, 313(1), pp. 27–34. doi: 10.1006/JMBI.2001.4969.

McEwen, K. R. and Ferguson-Smith, A. C. (2010) ‘Distinguishing epigenetic marks of developmental and imprinting regulation’, *Epigenetics & Chromatin*. BioMed Central, 3(1), p. 2. doi: 10.1186/1756-8935-3-2.

McGrath, J. and Solter, D. (1983) ‘Nuclear transplantation in mouse embryos’, *Journal of Experimental Zoology*. John Wiley & Sons, Ltd, 228(2), pp. 355–362. doi: 10.1002/jez.1402280218.

McGrath, J. and Solter, D. (1984a) ‘Completion of mouse embryogenesis requires both the maternal and paternal genomes’, *Cell*. Cell Press, 37(1), pp. 179–183. doi: 10.1016/0092-8674(84)90313-1.

McGrath, J. and Solter, D. (1984b) ‘Maternal Thp lethality in the mouse is a nuclear, not cytoplasmic, defect’, *Nature*. Nature Publishing Group, 308(5959), pp. 550–551. doi: 10.1038/308550a0.

McVey, M. and Lee, S. E. (2008) ‘MMEJ repair of double-strand breaks (director’s cut): deleted sequences and alternative endings’, *Trends in Genetics*. Elsevier Current Trends, 24(11), pp. 529–538. doi: 10.1016/J.TIG.2008.08.007.

Meehan, R. R. *et al.* (1989) ‘Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs’, *Cell*, 58(3), pp. 499–507. doi: 10.1016/0092-8674(89)90430-3.

Menke, D. B. (2013) ‘Engineering subtle targeted mutations into the mouse genome’, *genesis*. John Wiley & Sons, Ltd, 51(9), p. n/a-n/a. doi: 10.1002/dvg.22422.

Mersfelder, E. L. and Parthun, M. R. (2006) ‘The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure’, *Nucleic Acids Research*, 34(9), pp. 2653–2662. doi: 10.1093/nar/gkl338.

Meshorer, E. (2007) ‘Chromatin in embryonic stem cell neuronal differentiation’, *Histology and Histopathology*, 22(1–3), pp. 311–319. doi: 10.14670/HH-22.311.

Meuleman, W. *et al.* (2013) ‘Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence’, *Genome Research*. Cold Spring Harbor Laboratory Press, 23(2), pp. 270–280. doi: 10.1101/GR.141028.112.

Mikkelsen, T. S. *et al.* (2007) ‘Genome-wide maps of chromatin state in pluripotent and lineage-committed cells’, *Nature*. Nature Publishing Group, 448(7153), pp. 553–560. doi: 10.1038/nature06008.

Miller, J. C. *et al.* (2011) ‘A TALE nuclease architecture for efficient genome editing’, *Nature Biotechnology*. Nature Publishing Group, 29(2), pp. 143–148. doi: 10.1038/nbt.1755.

Mladenov, E. *et al.* (2013) ‘DNA Double-Strand Break Repair as Determinant of Cellular Radiosensitivity to Killing and Target in Radiation Therapy’, *Frontiers in Oncology*. Frontiers, 3, p. 113. doi: 10.3389/fonc.2013.00113.

Mladenov, E. *et al.* (2016) ‘DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: Cell cycle and proliferation-dependent regulation’, *Seminars in Cancer Biology*. Academic Press, 37–38, pp. 51–64. doi: 10.1016/J.SEMCANCER.2016.03.003.

Mohammad, F. *et al.* (2010) ‘Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1.’, *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 137(15), pp. 2493–9. doi: 10.1242/dev.048181.

Mohammad, F. *et al.* (2012) ‘Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing’, *Development*. Oxford University Press for The Company of Biologists Limited, 139(15), pp. 2792–2803. doi: 10.1242/DEV.079566.

Mojica, F. J. M. *et al.* (2009) ‘Short motif sequences determine the targets of

the prokaryotic CRISPR defence system', *Microbiology*. Microbiology Society, 155(3), pp. 733–740. doi: 10.1099/mic.0.023960-0.

Morbitzer, R. *et al.* (2010) 'Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(50), pp. 21617–22. doi: 10.1073/pnas.1013133107.

Moreno-Mateos, M. A. *et al.* (2015) 'CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo', *Nature Methods*. Nature Publishing Group, 12(10), pp. 982–988. doi: 10.1038/nmeth.3543.

Moshous, D. *et al.* (2001) 'Artemis, a Novel DNA Double-Strand Break Repair/V(D)J Recombination Protein, Is Mutated in Human Severe Combined Immune Deficiency', *Cell*. Cell Press, 105(2), pp. 177–186. doi: 10.1016/S0092-8674(01)00309-9.

Moynahan, M. E., Pierce, A. J. and Jasin, M. (2001) 'BRCA2 Is Required for Homology-Directed Repair of Chromosomal Breaks', *Molecular Cell*. Cell Press, 7(2), pp. 263–272. doi: 10.1016/S1097-2765(01)00174-5.

Muller, H. J. (1928) 'The Production of Mutations by X-Rays', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 14(9), p. 714. doi: 10.1073/PNAS.14.9.714.

Murphy, S. K. *et al.* (2015) 'Genotype-epigenotype interaction at the IGF2 DMR', *Genes*. MDPI AG, 6(3), pp. 777–789. doi: 10.3390/genes6030777.

Murrell, A., Heeson, S. and Reik, W. (2004) 'Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops', *Nature Genetics*. Nature Publishing Group, 36(8), pp. 889–893. doi: 10.1038/ng1402.

Myant, K. *et al.* (2011) 'LSH and G9a/GLP complex are required for developmentally programmed DNA methylation', *Genome Research*. Cold Spring Harbor Laboratory Press, 21(1), pp. 83–94. doi:

10.1101/gr.108498.110.

Nagano, T. *et al.* (2008) ‘The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 322(5908), pp. 1717–20. doi: 10.1126/science.1163802.

Nagy, A. *et al.* (2008) ‘Karyotyping mouse cells.’, *CSH protocols*. Cold Spring Harbor Laboratory Press, 2008(5), p. pdb.prot4706. doi: 10.1101/pdb.prot4706.

Nakamura, T. *et al.* (2007) ‘PGC7/Stella protects against DNA demethylation in early embryogenesis’, *Nature Cell Biology*. Nature Publishing Group, 9(1), pp. 64–71. doi: 10.1038/ncb1519.

Nambiar, T. S. *et al.* (2019) ‘Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant’, *Nature Communications*. Nature Publishing Group, 10(1), p. 3395. doi: 10.1038/s41467-019-11105-z.

Nan, X. *et al.* (1998) ‘Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex’, *Nature*. Nature Publishing Group, 393(6683), pp. 386–389. doi: 10.1038/30764.

Nelson, C. E. *et al.* (2019) ‘Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy’, *Nature Medicine*. Nature Publishing Group, 25(3), pp. 427–432. doi: 10.1038/s41591-019-0344-3.

Ng, H. H. *et al.* (1999) ‘MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex’, *Nature Genetics*. Presse Dienstleistungsgesellschaft mbH und Co. KG, 23(1), pp. 58–61. doi: 10.1038/12659.

O’Malley, R. C. *et al.* (2016) ‘Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape’, *Cell*. Cell Press, 165(5), pp. 1280–1292. doi: 10.1016/j.cell.2016.04.038.

Obata, Y. *et al.* (1998) ‘Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis’,

Development, 125(8).

Ochi, T. *et al.* (2015) ‘DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 347(6218), pp. 185–188. doi: 10.1126/science.1261971.

Okano, M. *et al.* (1999) ‘DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development’, *Cell*. Cell Press, 99(3), pp. 247–257. doi: 10.1016/S0092-8674(00)81656-6.

Okano, M., Xie, S. and Li, E. (1998) ‘Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases’, *Nature Genetics*. Nature Publishing Group, pp. 219–220. doi: 10.1038/890.

Okas, Y., Rozek, L. M. and Czech, M. P. (1985) ‘Direct Demonstration of Rapid Insulin-like Growth Factor I1 Receptor Internalization and Recycling in Rat Adipocytes INSULIN STIMULATES ’251-INSULIN-LIKE GROWTH FACTOR I1 DEGRADATION BY MODULATING THE IGF-I1 RECEPTOR RECYCLING PROCESS*’, *The Journal of Biological Chemistry*, 260(16), pp. 9435–9442.

Ooi, S. K. T. *et al.* (2007) ‘DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA’, *Nature*. Nature Publishing Group, 448(7154), pp. 714–717. doi: 10.1038/nature05987.

Orthwein, A. *et al.* (2015) ‘A mechanism for the suppression of homologous recombination in G1 cells’, *Nature*. Nature Publishing Group, 528(7582), pp. 422–426. doi: 10.1038/nature16142.

Otani, J. *et al.* (2013) ‘Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells’, *PLoS ONE*. Public Library of Science, 8(12). doi: 10.1371/journal.pone.0082961.

Pabo, C. O., Peisach, E. and Grant, R. A. (2001) ‘Design and Selection of Novel Cys $_2$ His $_2$ Zinc Finger Proteins’, *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA

94303-0139, USA , 70(1), pp. 313–340. doi: 10.1146/annurev.biochem.70.1.313.

Pandey, R. R. *et al.* (2008) ‘Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation’, *Molecular Cell*. Cell Press, 32(2), pp. 232–246. doi: 10.1016/J.MOLCEL.2008.08.022.

Pardue, M. L. and Gall, J. G. (1970) ‘Chromosomal localization of mouse satellite DNA.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 168(3937), pp. 1356–8. doi: 10.1126/science.168.3937.1356.

Paull, T. T. (2005) ‘Saving the Ends for Last: The Role of Pol μ in DNA End Joining’, *Molecular Cell*. Cell Press, 19(3), pp. 294–296. doi: 10.1016/J.MOLCEL.2005.07.008.

Paulsen, B. S. *et al.* (2017) ‘Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing’, *Nature Biomedical Engineering 2017 1:11*. Nature Publishing Group, 1(11), p. 878. doi: 10.1038/s41551-017-0145-2.

Pérez-González, J. A., Vara, J. and Jiménez, A. (1983) ‘Acetylation of puromycin by Streptomyces, alboniger the producing organism’, *Biochemical and Biophysical Research Communications*. Academic Press, 113(3), pp. 772–777. doi: 10.1016/0006-291X(83)91066-5.

Plagge, A. *et al.* (2004) ‘The imprinted signaling protein XLαs is required for postnatal adaptation to feeding’, *Nature Genetics*. Nature Publishing Group, 36(8), pp. 818–826. doi: 10.1038/ng1397.

Polstein, L. R. *et al.* (2015) ‘Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators’, *Genome Research*, 25(8). doi: 10.1101/gr.179044.114.

Porteus, M. H. and Baltimore, D. (2003) ‘Chimeric Nucleases Stimulate Gene

Targeting in Human Cells', *Science*, 300(5620), pp. 763–763. doi: 10.1126/science.1078395.

Prakash, R. *et al.* (2015) 'Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 7(4), p. a016600. doi: 10.1101/cshperspect.a016600.

Quenneville, S. *et al.* (2011) 'In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions', *Molecular Cell*. Cell Press, 44(3), pp. 361–372. doi: 10.1016/J.MOLCEL.2011.08.032.

Ramirez, C. L. *et al.* (2008) 'Unexpected failure rates for modular assembly of engineered zinc fingers', *Nature Methods*. Nature Publishing Group, 5(5), pp. 374–375. doi: 10.1038/nmeth0508-374.

Ran, F. A., Hsu, P. D., Lin, C.-Y., *et al.* (2013) 'Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.', *Cell*. Elsevier, 154(6), pp. 1380–9. doi: 10.1016/j.cell.2013.08.021.

Ran, F. A., Hsu, P. D., Wright, J., *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*. Nature Publishing Group, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Regha, K. *et al.* (2007) 'Active and Repressive Chromatin Are Interspersed without Spreading in an Imprinted Gene Cluster in the Mammalian Genome', *Molecular Cell*. Cell Press, 27(3), pp. 353–366. doi: 10.1016/J.MOLCEL.2007.06.024.

Reik, W., Dean, W. and Walter, J. (2001) 'Epigenetic reprogramming in mammalian development.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 293(5532), pp. 1089–93. doi: 10.1126/science.1063443.

Reik, W. and Walter, J. (2001) 'Genomic imprinting: parental influence on the genome', *Nature Reviews Genetics*. Nature Publishing Group, 2(1), pp. 21–

32. doi: 10.1038/35047554.

Richardson, C. D. *et al.* (2016) ‘Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA’, *Nature Biotechnology*. Nature Publishing Group, 34(3), pp. 339–344. doi: 10.1038/nbt.3481.

Richardson, C. D. *et al.* (2018) ‘CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway’, *Nature Genetics*. Nature Publishing Group, 50(8), pp. 1132–1139. doi: 10.1038/s41588-018-0174-0.

Roberts, S. A. *et al.* (2010) ‘Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends’, *Nature*. Nature Publishing Group, 464(7292), pp. 1214–1217. doi: 10.1038/nature08926.

Robertson, K. D. (2005) ‘DNA methylation and human disease’, *Nature Reviews Genetics*. Nature Publishing Group, 6(8), pp. 597–610. doi: 10.1038/nrg1655.

Rosidi, B. *et al.* (2008) ‘Histone H1 functions as a stimulatory factor in backup pathways of NHEJ’, *Nucleic Acids Research*, 36(5), pp. 1610–1623. doi: 10.1093/nar/gkn013.

Rouet, P., Smih, F. and Jasin, M. (1994) ‘Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 91(13), pp. 6064–8. doi: 10.1073/pnas.91.13.6064.

Rowe, H. M. *et al.* (2010) ‘KAP1 controls endogenous retroviruses in embryonic stem cells’, *Nature*. Nature Publishing Group, 463(7278), pp. 237–240. doi: 10.1038/nature08674.

Roy, S. *et al.* (2015) ‘XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation.’, *Molecular and cellular biology*. American Society for Microbiology Journals, 35(17), pp. 3017–28. doi: 10.1128/MCB.01503-14.

Rudin, N., Sugarman, E. and Haber, J. E. (1989) ‘Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae.’, *Genetics*, 122(3).

San Filippo, J., Sung, P. and Klein, H. (2008) ‘Mechanism of Eukaryotic Homologous Recombination’, *Annual Review of Biochemistry*. Annual Reviews , 77(1), pp. 229–257. doi: 10.1146/annurev.biochem.77.061306.125255.

Sanjana, N. E. *et al.* (2012) ‘A transcription activator-like effector toolbox for genome engineering’, *Nature Protocols*. Nature Publishing Group, 7(1), pp. 171–192. doi: 10.1038/nprot.2011.431.

Sanli, I. *et al.* (2018) ‘Meg3 Non-coding RNA Expression Controls Imprinting by Preventing Transcriptional Upregulation in cis’, *Cell Reports*. Cell Press, 23(2), pp. 337–348. doi: 10.1016/J.CELREP.2018.03.044.

Schatz, D. G. and Swanson, P. C. (2011) ‘V(D)J Recombination: Mechanisms of Initiation’, *Annual Review of Genetics*. Annual Reviews , 45(1), pp. 167–202. doi: 10.1146/annurev-genet-110410-132552.

Schulz, R. *et al.* (2010) ‘The Parental Non-Equivalence of Imprinting Control Regions during Mammalian Development and Evolution’, *PLoS Genetics*. Edited by W. Reik. Public Library of Science, 6(11), p. e1001214. doi: 10.1371/journal.pgen.1001214.

Shalem, O. *et al.* (2014) ‘Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells’, *Science*, 343(6166), pp. 84–87. doi: 10.1126/science.1247005.

Sharif, J. *et al.* (2007) ‘The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA’, *Nature*. Nature Publishing Group, 450(7171), pp. 908–912. doi: 10.1038/nature06397.

Shen, B. *et al.* (2014) ‘Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects’, *Nature Methods*. Nature Publishing Group, 11(4), pp. 399–402. doi: 10.1038/nmeth.2857.

Shen, M. W. *et al.* (2018) ‘Predictable and precise template-free CRISPR editing of pathogenic variants’, *Nature*. Nature Publishing Group, 563(7733), pp. 646–651. doi: 10.1038/s41586-018-0686-x.

Shou, J. *et al.* (2018) ‘Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion’, *Molecular Cell*. Cell Press, 71(4), pp. 498-509.e4. doi: 10.1016/j.molcel.2018.06.021.

Sleutels, F., Zwart, R. and Barlow, D. P. (2002) ‘The non-coding Air RNA is required for silencing autosomal imprinted genes’, *Nature*. Nature Publishing Group, 415(6873), pp. 810–813. doi: 10.1038/415810a.

Smilinich, N. J. *et al.* (1999) ‘A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(14), pp. 8064–9. doi: 10.1073/pnas.96.14.8064.

Smithies, O. *et al.* (1985) ‘Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination’, *Nature*. Nature Publishing Group, 317(6034), pp. 230–234. doi: 10.1038/317230a0.

Stanford, W. L., Cohn, J. B. and Cordes, S. P. (2001) ‘Gene-trap mutagenesis: past, present and beyond’, *Nature Reviews Genetics*. Nature Publishing Group, 2(10), pp. 756–768. doi: 10.1038/35093548.

Stark, J. M. and Jasin, M. (2003) ‘Extensive Loss of Heterozygosity Is Suppressed during Homologous Repair of Chromosomal Breaks’, *Molecular and Cellular Biology*, 23(2), pp. 733–743. doi: 10.1128/MCB.23.2.733-743.2003.

van Steensel, B. and Belmont, A. S. (2017) ‘Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression’, *Cell*. Cell Press, 169(5), pp. 780–791. doi: 10.1016/J.CELL.2017.04.022.

Sternberg, S. H. *et al.* (2014) ‘DNA interrogation by the CRISPR RNA-guided endonuclease Cas9’, *Nature*. Nature Publishing Group, 507(7490), pp. 62–67. doi: 10.1038/nature13011.

Strande, N. *et al.* (2012) ‘Specificity of the dRP/AP lyase of Ku promotes nonhomologous end joining (NHEJ) fidelity at damaged ends.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 287(17), pp. 13686–93. doi: 10.1074/jbc.M111.329730.

Straub, T. (2003) ‘Heterochromatin Dynamics’, *PLoS Biology*. Public Library of Science, 1(1), p. e14. doi: 10.1371/journal.pbio.0000014.

Surani, M. A. and Barton, S. C. (1983) ‘Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 222(4627), pp. 1034–6. doi: 10.1126/science.6648518.

Surani, M. A. H., Barton, S. C. and Norris, M. L. (1984) ‘Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis’, *Nature*. Nature Publishing Group, 308(5959), pp. 548–550. doi: 10.1038/308548a0.

Symington, L. S. (2002) ‘Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair.’, *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 66(4), pp. 630–70, table of contents. doi: 10.1128/mmbr.66.4.630-670.2002.

Szabó, P. E. *et al.* (2000) ‘Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function’, *Current Biology*. Cell Press, 10(10), pp. 607–610. doi: 10.1016/S0960-9822(00)00489-9.

Tahiliani, M. *et al.* (2009) ‘Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1’, *Science*. American Association for the Advancement of Science, 324(5929), pp. 930–935. doi: 10.1126/science.1170116.

Takada, T. *et al.* (2013) ‘The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains.’, *Genome research*. Cold Spring Harbor Laboratory Press, 23(8), pp. 1329–38. doi: 10.1101/gr.156497.113.

Tan, W. S. *et al.* (2012) ‘Precision editing of large animal genomes.’, *Advances in genetics*. NIH Public Access, 80, pp. 37–97. doi: 10.1016/B978-0-12-404742-6.00002-8.

Tao, Y. *et al.* (2011) ‘Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(14), pp. 5626–5631. doi: 10.1073/pnas.1017000108.

Terranova, R. *et al.* (2008) ‘Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos’, *Developmental Cell*. Cell Press, 15(5), pp. 668–679. doi: 10.1016/J.DEVCEL.2008.08.015.

Thomas, K. R. and Capecchi, M. R. (1987) ‘Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells’, *Cell*. Cell Press, 51(3), pp. 503–512. doi: 10.1016/0092-8674(87)90646-5.

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986) ‘High frequency targeting of genes to specific sites in the mammalian genome’, *Cell*. Cell Press, 44(3), pp. 419–428. doi: 10.1016/0092-8674(86)90463-0.

Tokunaga, A., Anai, H. and Hanada, K. (2016) ‘Mechanisms of gene targeting in higher eukaryotes’, *Cellular and Molecular Life Sciences*. Springer International Publishing, 73(3), pp. 523–533. doi: 10.1007/s00018-015-2073-1.

Torres, R. *et al.* (2014) ‘Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system’, *Nature Communications*. Nature Publishing Group, 5(1), p. 3964. doi: 10.1038/ncomms4964.

Tóth, K. F. *et al.* (2004) ‘Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin’, *Journal of Cell Science*. The Company of Biologists Ltd, 117(18), pp. 4277–4287. doi: 10.1242/JCS.01293.

Trojer, P. and Reinberg, D. (2007) ‘Facultative Heterochromatin: Is There a Distinctive Molecular Signature?’, *Molecular Cell*. Cell Press, 28(1), pp. 1–13. doi: 10.1016/J.MOLCEL.2007.09.011.

Tsai, S. Q. *et al.* (2015) ‘GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases’, *Nature Biotechnology*. Nature Publishing Group, 33(2), pp. 187–197. doi: 10.1038/nbt.3117.

Tsai, S. Q. *et al.* (2017) ‘CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets’, *Nature Methods*. Nature Publishing Group, 14(6), pp. 607–614. doi: 10.1038/nmeth.4278.

Tsouroula, K. *et al.* (2016) ‘Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin’, *Molecular Cell*. Elsevier Inc., 63(2), pp. 293–305. doi: 10.1016/j.molcel.2016.06.002.

Ueda, T. *et al.* (2000) ‘The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development’, *Genes to Cells*. John Wiley & Sons, Ltd (10.1111), 5(8), pp. 649–659. doi: 10.1046/j.1365-2443.2000.00351.x.

Urnov, F. D. *et al.* (2005) ‘Highly efficient endogenous human gene correction using designed zinc-finger nucleases’, *Nature*. Nature Publishing Group, 435(7042), pp. 646–651. doi: 10.1038/nature03556.

Urnov, F. D. *et al.* (2010) ‘Genome editing with engineered zinc finger nucleases’, *Nature Reviews Genetics*. Nature Publishing Group, 11(9), pp. 636–646. doi: 10.1038/nrg2842.

Uusi-Mäkelä, M. I. E. *et al.* (2018) ‘Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio)’, *PLOS ONE*. Edited

by B. B. Riley. Public Library of Science, 13(4), p. e0196238. doi: 10.1371/journal.pone.0196238.

van Overbeek, M. *et al.* (2016) ‘DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks’, *Molecular Cell*. Cell Press, 63(4), pp. 633–646. doi: 10.1016/J.MOLCEL.2016.06.037.

Verkuijl, S. A. and Rots, M. G. (2019) ‘The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies’, *Current Opinion in Biotechnology*. Elsevier Current Trends, 55, pp. 68–73. doi: 10.1016/J.COPBIO.2018.07.005.

Villarreal, D. D. *et al.* (2012) ‘Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations’, *PLoS Genetics*. Edited by M. Lichten. Public Library of Science, 8(11), p. e1003026. doi: 10.1371/journal.pgen.1003026.

Vostrov, A. A. and Quitschke, W. W. (1997) ‘The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 272(52), pp. 33353–9. doi: 10.1074/jbc.272.52.33353.

Walter, M. *et al.* (2016) ‘An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells’, *eLife*, 5. doi: 10.7554/eLife.11418.

Wang, M. *et al.* (2006) ‘PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways’, *Nucleic Acids Research*, 34(21), pp. 6170–6182. doi: 10.1093/nar/gkl840.

Wang, T. *et al.* (2014) ‘Genetic screens in human cells using the CRISPR-Cas9 system.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 343(6166), pp. 80–4. doi: 10.1126/science.1246981.

Wang, Z. *et al.* (2008) ‘Combinatorial patterns of histone acetylations and

methylations in the human genome', *Nature Genetics*. Nature Publishing Group, 40(7), pp. 897–903. doi: 10.1038/ng.154.

Wang, Z. *et al.* (2009) 'Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes', *Cell*. Elsevier, 138(5), pp. 1019–1031. doi: 10.1016/J.CELL.2009.06.049.

Weaver, J. R. *et al.* (2010) 'Domain-specific response of imprinted genes to reduced DNMT1.', *Molecular and cellular biology*. American Society for Microbiology Journals, 30(16), pp. 3916–28. doi: 10.1128/MCB.01278-09.

Weber, A. R. *et al.* (2016) 'Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism', *Nature Communications*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/ncomms10806.

Weber, M. *et al.* (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nature Genetics*. Nature Publishing Group, 39(4), pp. 457–466. doi: 10.1038/ng1990.

White, J. and Dalton, S. (2005) 'Cell Cycle Control of Embryonic Stem Cells', *Stem Cell Reviews*. Humana Press, 1(2), pp. 131–138. doi: 10.1385/SCR:1:2:131.

Wienert, B. *et al.* (2019) 'Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq', *Science*. American Association for the Advancement of Science, 364(6437), pp. 286–289. doi: 10.1126/SCIENCE.AAV9023.

Williams, G. J., Lees-Miller, S. P. and Tainer, J. A. (2010) 'Mre11–Rad50–Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks', *DNA Repair*. Elsevier, 9(12), pp. 1299–1306. doi: 10.1016/J.DNAREP.2010.10.001.

Williams, R. S. *et al.* (2008) 'Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair', *Cell*. Cell Press, 135(1), pp. 97–109. doi: 10.1016/J.CELL.2008.08.017.

Wood, A. J. *et al.* (2008) ‘Regulation of alternative polyadenylation by genomic imprinting.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 22(9), pp. 1141–6. doi: 10.1101/gad.473408.

Woodcock, C. L. and Dimitrov, S. (2001) ‘Higher-order structure of chromatin and chromosomes’, *Current Opinion in Genetics & Development*. Elsevier Current Trends, 11(2), pp. 130–135. doi: 10.1016/S0959-437X(00)00169-6.

Woodcock, C. L. and Ghosh, R. P. (2010) ‘Chromatin higher-order structure and dynamics.’, *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 2(5), p. a000596. doi: 10.1101/cshperspect.a000596.

Wray, J. *et al.* (2011) ‘Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation’, *Nature Cell Biology*. Nature Publishing Group, 13(7), pp. 838–845. doi: 10.1038/ncb2267.

Wu, X. *et al.* (2014) ‘Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells’, *Nature Biotechnology*. Nature Publishing Group, 32(7), pp. 670–676. doi: 10.1038/nbt.2889.

Wutz, A. *et al.* (1997) ‘Imprinted expression of the Igf2r gene depends on an intronic CpG island’, *Nature*. Nature Publishing Group, 389(6652), pp. 745–749. doi: 10.1038/39631.

Wutz, A. *et al.* (2001) ‘Non-imprinted Igf2r expression decreases growth and rescues the Tme mutation in mice.’, *Development (Cambridge, England)*. The Company of Biologists Ltd, 128(10), pp. 1881–7. doi: 10.1242/dev.032060.

Xia, B. *et al.* (2006) ‘Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2’, *Molecular Cell*. Cell Press, 22(6), pp. 719–729. doi: 10.1016/J.MOLCEL.2006.05.022.

Xu, H. *et al.* (2012) ‘FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads’, *PLoS ONE*. Edited by D. Doucet. Public Library of

Science, 7(12), p. e52249. doi: 10.1371/journal.pone.0052249.

Yagi, M. *et al.* (2017) ‘Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation’, *Nature*. Nature Publishing Group, 548(7666), pp. 224–227. doi: 10.1038/nature23286.

Yáñez, R. and Porter, A. (1998) ‘Therapeutic gene targeting’, *Gene Therapy*. Nature Publishing Group, 5(2), pp. 149–159. doi: 10.1038/sj.gt.3300601.

Yang, Y. *et al.* (2016) ‘A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice’, *Nature Biotechnology*. Nature Publishing Group, 34(3), pp. 334–338. doi: 10.1038/nbt.3469.

Yarrington, R. M. *et al.* (2018) ‘Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(38), pp. 9351–9358. doi: 10.1073/pnas.1810062115.

Yin, Y. *et al.* (2017) ‘Impact of cytosine methylation on DNA binding specificities of human transcription factors’, *Science*. American Association for the Advancement of Science, 356(6337). doi: 10.1126/science.aaj2239.

Yoon, H. G. *et al.* (2003) ‘N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso’, *Molecular Cell*. Cell Press, 12(3), pp. 723–734. doi: 10.1016/j.molcel.2003.08.008.

Yusufzai, T. M. *et al.* (2004) ‘CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species’, *Molecular Cell*. Cell Press, 13(2), pp. 291–298. doi: 10.1016/S1097-2765(04)00029-2.

Zakharyevich, K. *et al.* (2010) ‘Temporally and Biochemically Distinct Activities of Exo1 during Meiosis: Double-Strand Break Resection and Resolution of Double Holliday Junctions’, *Molecular Cell*. Cell Press, 40(6), pp. 1001–1015. doi: 10.1016/J.MOLCEL.2010.11.032.

Zavorka, M. E. *et al.* (2016) ‘Inhibition of insulin-like growth factor II (IGF-II)-dependent cell growth by multidentate pentamannosyl 6-phosphate-based

ligands targeting the mannose 6-phosphate/IGF-II receptor', *Oncotarget*. Impact Journals LLC, 7(38), pp. 62386–62410. doi: 10.18632/oncotarget.11493.

Zemach, A. *et al.* (2010) 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*. American Association for the Advancement of Science, 328(5980), pp. 916–919. doi: 10.1126/science.1186366.

Zhang, F. *et al.* (2011) 'Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription', *Nature Biotechnology*. Nature Publishing Group, 29(2), pp. 149–153. doi: 10.1038/nbt.1775.

Zvetkova, I. *et al.* (2005) 'Global hypomethylation of the genome in XX embryonic stem cells', *Nature Genetics*. Nature Publishing Group, 37(11), pp. 1274–1279. doi: 10.1038/ng1663.

Zwart, R. *et al.* (2001) 'Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes.', *Genes & development*. Cold Spring Harbor Laboratory Press, 15(18), pp. 2361–6. doi: 10.1101/gad.206201.

Appendices

Appendix 1 - Sequences of sgRNAs and ssODNs

1.1 Protospacer sequences and primers for sgRNAs

1.1.1 Protospacer sequences

sgRNA	Protospacer sequence
sgKvDMR#1	GTGCGTTCTGACTCGGCCCG
sgKvDMR#2	CTACCAAGACCTTGGTGACG
sgKvDMR#3	CTGTCCAATCAACAGTGTCG
sgImpact	GAAAGTGTTGGGAGTGACGT
sgInpp5f_v2	CTGCGGAGGATAGATGACCG
sgNcaph	GCTGCTGTCTTCTGAAACAT
sgGrb10_3´1	GAGGAACGGACGGTAAGACG
sgImpact_3´1	GTGTGAGTTTACAGGTCTGT
sgKvDMR#1 mm1	GTGCGTTCTGACTCGGCCCT
sgKvDMR#1 mm14	GTGCGTACTGACTCGGCCCG
sgKvDMR#1 mm1+10	GTGCGTTCTGGCTCGGCCCT
sgKvDMR#1 mm11+12	GTGCGTTCCAACTCGGCCCG
sgKvDMR#1 mm14+11	GTGCGTACTCACTCGGCCCG
sgKvDMR#1 mm14+11+8	GTGCGTACTCACCGGCCCG
sgImpact mm1	GAAAGTGTTGGGAGTGACGG
sgImpact mm14	GAAAGTCTTGGGAGTGACGT
sgImpact mm1+10	GAAAGTGTTGTGAGTGACGG
sgImpact mm11+12	GAAAGTGTCAGGAGTGACGT
sgImpact mm14+11	GAAAGTCTTAGGAGTGACGT
sgImpact mm14+11+8	GAAAGTCTTAGGTGTGACGT

1.1.2 Primers for sgRNAs

Primer Name	Primer Sequence
sgKvDMR#1_Fwd	CACCGTGCGTTCTGACTCGGCCCG
sgKvDMR#1_Rev	AAACCGGGCCGAGTCAGAACGCAC
sgKvDMR#2_Fwd	CACCGGCTACCAAGACCTTGGTGACG
sgKvDMR#2_Rev	AAACCGTCACCAAGGTCTTGGTAGCC
sgKvDMR#3_Fwd	CACCGGCTGTCCAATCAACAGTGTCG
sgKvDMR#3_Rev	AAACCGACACTGTTGATTGGACAGCC
sgImpact_Fwd	CACCGGAAAGTGTTGGGAGTGACGT
sgImpact_Rev	AAACACGTCACTCCCAACACTTTCC
sgInpp5f_v2_Fwd	CACCGCTGCGGAGGATAGATGACCG
sgInpp5f_v2_Rev	AAACCGGTCATCTATCCTCCGCAGC
sgNcaph_Fwd	CACCGCTGCTGTCTTCTGAAACAT
sgNcaph_Rev	AAACATGTTTCAGAAGACAGCAGC
sgGrb10_3´ 1_Fwd	CACCGAGGAACGGACGGTAAGACG
sgGrb10_3´ 1_Rev	AAACCGTCTTACCGTCCGTTCCTC
sgImpact_3´ 1_Fwd	CACCGTGTGAGTTTACAGGTCTGT
sgImpact_3´ 1_Rev	AAACACAGACCTGTAAACTCACAC
sgKvDMR#1 mm1_Fwd	CACCGTGCGTTCTGACTCGGCCCT
sgKvDMR#1 mm1_Rev	AAACAGTGCCGAGTCAGAACGCAC
sgKvDMR#1 mm14_Fwd	CACCGTGCGTACTGACTCGGCCCG
sgKvDMR#1 mm14_Rev	AAACCGGGCCGAGTCAGTACGCAC
sgKvDMR#1 mm1+10_Fwd	CACCGTGCGTTCTGGCTCGGCCCT
sgKvDMR#1 mm1+10_Rev	AAACAGGGCCGAGCCAGAACGCAC

sgKvDMR#1 mm11+12_Fwd	CACCGTGCGTTCCAACTCGGCCCG
sgKvDMR#1 mm11+12_Rev	AAACCGGGCCGAGTTGGAACGCAC
sgKvDMR#1 mm14+11_Fwd	CACCGTGCGTACTCACTCGGCCCG
sgKvDMR#1 mm14+11_Rev	AAACCGGGCCGAGTGAGTACGCAC
sgKvDMR#1 mm14+11+8_Fwd	CACCGTGCGTACTCACCCGGCCCG
sgKvDMR#1 mm14+11+8_Rev	AAACCGGGCCGGGTGAGTACGCAC
sgImpact mm1_Fwd	CACCGAAAGTGTTGGGAGTGACGG
sgImpact mm1_Rev	AAACCCGTCACTCCCAACACTTTC
sgImpact mm14_Fwd	CACCGAAAGTCTTGGGAGTGACGT
sgImpact mm14_Rev	AAACACGTCACTCCCAAGACTTTC
sgImpact mm1+10_Fwd	CACCGAAAGTGTTGTGAGTGACGG
sgImpact mm1+1_Rev	AAACCCGTCACTCACAACACTTTC
sgImpact mm11+12_Fwd	CACCGAAAGTGTCAGGAGTGACGT
sgImpact mm11+12_Rev	AAACACGTCACTCCTGACACTTTC
sgImpact mm14+11_Fwd	CACCGAAAGTCTTAGGAGTGACGT
sgImpact mm14+11_Rev	AAACACGTCACTCCTAAGACTTTC
sgImpact mm14+11+8_Fwd	CACCGAAAGTCTTAGGTGTGACGT
sgImpact mm14+11+8_Rev	AAACACGTCACACCTAAGACTTTC

1.2 Single-stranded oligodinucleotide sequences

ssKvDMR#1

GGGCTACAAAGCTCAGGGGTCTCCAGACCCGATTCGGTTTCAGC
TCCAGTGCGTTCTGACTCGGCCCAATATTTAGAATCCGAAGGCCTGAGC
CGGTGTCCTAGGCCACTCACCTTGGGACTCGACCGACCTCGGGGCTCA
AAGGGCCTC

ssKvDMR#2&3

CAGACGTGCTGAGGCAACTGTTCCTCCTAGCGACAACGGCTAGG
CCACCTACCAAGACCTTGGTGACGTGGCAGCTGCGACACTGTTGATTG
GACAGATGCTGAATAATGACTAGGGAGCTCTTCCTCATGATAGGCTGGA
TACAGATAG

ssImpact

GGAGCCTCAGTGCGGCCTGCAAAAGGCCCTAGCCCCATGACTTT
GCAAGAAACAGAACCCAGGGCAAGCTCCCGGGACGTCACTCCCAACAC
TTTCAGAAAGCCTGTAGCTGGCAGCCGCTGCCTGCACGCAGGGTTGTT
GACCCGGAAG

ssInpp5f_v2

AAAGAGGCTTAGGTTCGAAAGGAACTAGGCGAGCTGCATCAGGA
CCCAGAGCACCTGCAGAAACGCCGACTCTCGGCCGGTCATCTATCCTC
CGCAGGGTATTGCGACTATCAGGATCGCGGCAAAGATAACAGCTTGGC
AGCCCTCTGC

ssNcaph

CTACCGCCGTCATTCTTTGGTTCCCTGTTTTCCTCGCGTGCAGGTGGATCTTATGTTTCAGAAGACAGCAGCCTCCTTTGATGAATGCAGCACGACTGGGG

Appendix 2 - Other media recipes

Custom 2i media recipe for mESCs from Heidi Mjøseng

Serum free ES medium (SFES) - 500 ml	
Substance	Amount
Neurobasal	250 ml
DMEM/F12	250 ml
BSA (7.5%)	3.33 ml
Pen/Strep	5 ml
N-2 supplement	2.5 ml
B27+RA	5 ml

Complete "2i" medium - 50 ml (max shelf life 3 days)		
Substance	Amount	
SFES (above stock)	50 ml	
PD0325901 [1mM stock]	50 ul	[1uM final]
CHIR99021 [3mM stock]	50 ul	[3uM final]
Glutamine [200mM stock]	500 ul	[2mM final]
LIF [10^6U/ml]	50 ul	1000U/ml
Monothioglycerol [11.9M stock]	0.63 ul	[1.5x10-4M final]

Appendix 3 - Illumina MiSeq primers

3.1 First round of amplification primers for allelic specificity amplicons

Primer	Sequence
KvDMR#1 Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNACACGTACTCCACTCACTACC
KvDMR#1 Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNTGGTCTTGAGGCCCTTTGAG
KvDMR#2&3 Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNCTGACTGGACCAAAATGCAC
KvDMR#2&3 Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNTGGCCAATAAAAATAGTCAGCA
Impact Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNAGGATGAGTATGAGAGTCCCA
Impact Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCACCCACTTCCTGAATTGCT
Inpp5fv2 Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNTGGAAAGAGGCTTAGGTTCGA
Inpp5fv2 Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNTCCACGGAGAATAAAGACCCC
Ncaph Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNTCCACACTTATCCATCTGAGCA
Ncaph Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCCCAGTCGTGCTGCATTCAT

3.2 First round of amplification primers for bisulfite-converted amplicons

Primer	Sequence
KvDMR Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNGAGAAAAGTATATTAAGGTGTATTAGAT
KvDMR Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNTATATCCAACCTATCATAAAAAAAA
Impact Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNTTTTAATGGTATTGGAGTTTTAGTG
Impact Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNAACAAACAACAAATAAATACAACTAC
Inpp5fv2 Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNTTAGGATTTAGAGTATTTGTAGAAA
Inpp5fv2 Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNACCCCACTAACACTTTAACCATAAAT

3.3 Second round of amplification primers for all amplicons

3.3.1 Forward primers

Index 1 Primers	Sequence
N_index1_N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG
N_index1_N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG
N_index1_N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG
N_index1_N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG
N_index1_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG
N_index1_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
N_index1_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG
N_index1_N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG
N_index1_N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG
N_index1_N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG
N_index1_N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG
N_index1_N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGG
N_index1_N708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGG
N_index1_N709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGG
N_index1_N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGG
N_index1_N717	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGG

3.3.2 Reverse primers

Index 2 Primers	Sequence
N_index2_S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
N_index2_S503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
N_index2_S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
N_index2_S506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
N_index2_S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
N_index2_S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC
N_index2_S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTC
N_index2_S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTC
N_index2_S513	AATGATACGGCGACCACCGAGATCTACACTCGACTAGTCGTCGGCAGCGTC
N_index2_S515	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCGTCGGCAGCGTC
N_index2_S516	AATGATACGGCGACCACCGAGATCTACACCCTAGAGTTCGTCGGCAGCGTC
N_index2_S517	AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGCAGCGTC

Appendix 4 - Primers for other assays

4.1 Primers for qPCR

qPCR Primer	Sequence
KvDMR#1_Fwd	TTAGAATCCGAAGGCCTGA
KvDMR#1_Rev	TCCGTATTCGTTTTGCAGTT
KvDMR#2&3_Fwd	ACGTGCTGAGGCAACTGTTCC
KvDMR#2&3_Rev	TCCAGCCTATCATGAGGAAGAGC
Impact_Fwd	ATTGGAGCCTCAGTGCG
Impact_Rev	TGCTTCCGGGTCAACAA
Inpp5f_v2_Fwd	CAGACACCGGATAGAAGCAG
Inpp5f_v2_Rev	TCCTGATAGTCGCAATACCC
IAP_Fwd	ATGAGCAAGCCACAAAGGAG
IAP_Rev	GAGCGGTTCTGAGATTGGAG
Actb_Fwd	CCTCGATGCTGACCCTCATCC
Actb_Rev	GACACTGCCCCATTCAATGTCTC

4.2 Primers for Sanger sequencing

Sanger Primer	Sequence
KvDMR#1_Fwd	CAGCCTCAGTTCCACGATAC
KvDMR#1_Rev	AGAAGCAGAGGTGATTCGTG
KvDMR#2&3_Fwd	TGCACCATCATAGACCACGC
KvDMR#2&3_Rev	TGGAGACATCAGGCTGTTTCC
Impact_Fwd	TGAAGGATGAGTATGAGAGTCCC
Impact_Rev	AGACGACTGGCAAGCAGC
Inpp5f_v2_Fwd	ATAGAAGCAGCCCTCCTCAA
Inpp5f_v2_Rev	GAGAACGCCATCGGACTTAG

4.3 Primers for the Surveyor assay

Surveyor Primer	Sequence
KvDMR#1_Fwd	CGGAACCACTGTAGACCCAC
KvDMR#1_Rev	GTTTTGCAGTTACGGCTCGG
KvDMR#2&3_Fwd	CAGACGTGCTGAGGCAACTG
KvDMR#2&3_Rev	TGGCCTCCTTGTGTGCTTTAG
Impact_Fwd	CCTCTCCAGCTCTCGTTCATTT
Impact_Rev	ATCAAAAGGAAAGGCGGGGAA
Inpp5f_v2_Fwd	GAAGCGGTATTGGGAACTGTC
Inpp5f_v2_Rev	TCTGTCTACACCGCACCTCG
Ncaph_Fwd	TGGGTGAGCTTGTGTCTCCTG
Ncaph_Rev	TTACTGCAGAAGGTCCACAGCA
Grb10_3´_Fwd	CTCGCGGGAGATACGTCCAT
Grb10_3´_Rev	TTAGGGGACAATGCTCAGCC
Impact_3´_Fwd	GGTACTGAGGTGTGGACTAGGA
Impact_3´_Rev	AATGAAGCACCACAGCCATGT

Appendix 5 - Buffers and solutions

5.1 Buffers for DNase I accessibility assay

Buffer A:

- 15 mM Tris HCl (pH 7.6)
- 60 mM KCl, 15 mM NaCl
- 1 mM EDTA
- 0.5 mM EGTA
- 0.5 mM spermidine
- 0.15 mM spermine

Digestion buffer: buffer A supplemented with

- 3 mM CaCl2
- 75 mM NaCl

Stop buffer:

- 0.1 M NaCl
- 0.1 % (w/v) SDS
- 50 mM Tris-HCl (pH 8.0)
- 100 mM EDTA

TE buffer:

- 10 mM Tris-HCl (pH 8.0)
- 1 mM EDTA

5.2 Buffers for Native histone ChIP

NBA buffer:

- 85mM NaCl
- 5.5% sucrose
- 10mM Tris-HCl pH 7.5
- 0.2mM EDTA
- 0.2mM PMSF
- 1mM DTT
- protease inhibitors (Merck Millipore, 539131)

NBB buffer: NBA buffer with

- 0.1% NP-40

NBR buffer:

- 85mM NaCl
- 5.5% sucrose
- 10mM Tris-HCl (pH 7.5)
- 3mM MgCl2
- 1.5mM CaCl2
- 0.2mM PMSF
- 1mM DTT)

MNase stop buffer:

- 215mM NaCl
- 10mM TrisHCl pH8
- 20mM EDTA
- 5.5% sucrose
- 2% TritonX100
- 0.2mM PMSF
- 1mM DTT
- 2x PMSF

Blocking buffer:

- 0.5% BSA in PBS

ChIP-W1 buffer:

- 150mM NaCl
- 10mM Tris HCl (pH 8.0)
- 2mM EDTA
- 1% NP40#
- 1% Sodium Deoxycholate

TE buffer:

- 10 mM Tris-HCl (pH 8.0)
- 1 mM EDTA)

Elution buffer:

- 0.1mM NaHCO3
- 1% SDS

- **5.3 Buffers for cross-linked ChIP**

Farnham lysis buffer:

- 5mM PIPES (pH 8.0)
- 85mM KCl
- 0.5% NP-40
- protease inhibitors (Merck Millipore, 539131)

RIPA buffer:

- 1x PBS
- 1% NP-40
- 0.5% sodium deoxycholate
- 0.1% SDS
- protease inhibitors (Merck Millipore, 539131)

LiCl Wash Buffer:

- 100mM Tris (pH 7.5)
- 500mM LiCl
- 1%NP-40
- 0.5% sodium deoxycholate

Elution buffer:

- 0.1mM NaHCO3
- 1% SDS

Appendix 6 – Standard deviation of InDel classes

This is relevant to the data shown in Figure 5.5 and Figure 5.6.

- **sgKvDMR#1**

BxJ maternal	STDEV	BxJ paternal	STDEV
Δ5 (combined)	0.029383631	Δ5 (combined)	0.01684022
Δ1	0.01824212	Δ1	0.010263808
Δ2	0.016073368	Δ2	0.022510046
I1	0.003327726	I1	0.013761138

JxB maternal	STDEV	JxB paternal	STDEV
Δ5 (combined)	0.017277037	Δ5 (combined)	0.026910128
Δ1	0.042613555	Δ1	0.010887627
Δ2	0.018858367	Δ2	0.030471327
Δ7	0.020236899	Δ7	0.006330598

- **sgKvDMR#3**

BxJ maternal	STDEV	BxJ paternal	STDEV
Δ5	0.101980135	Δ5	0.01989343
Δ2	0.056707048	Δ2	0.01268864
Δ6	0.011177243	I1	0.02075783
I1	0.013393223	Δ6	0.02384817
Δ16	0.010965797	Δ1	0.01934358

JxB maternal	STDEV	JxB paternal	STDEV
Δ5	0.080880029	Δ5	0.03324798
Δ2	0.029249498	Δ2	0.01790599
Δ6	0.049424894	I1	0.01390138
I1	0.083117846	Δ6	0.01487232
Δ16	0.004243095	Δ1	0.01227614

- **sgInpp5f_v2**

BxJ maternal	**STDEV**	**BxJ paternal**	**STDEV**
Δ5	0.008172038	Δ5	0.005008938
Δ1	0.008372349	Δ1	0.015668178
Δ3	0.010804898	Δ3	0.007720155
Δ2	0.013545753	Δ6	0.00808854
Δ21	0.002074258	Δ13	0.031161081

JxB maternal	**STDEV**	**JxB paternal**	**STDEV**
Δ5	0.005898988	Δ5	0.011401644
Δ1	0.017217324	Δ1	0.006058997
Δ3	0.016047504	Δ3	0.011046054
Δ6	0.007282596	Δ2	0.016215758
Δ14	0.00268511	Δ6	0.008775066

- **sgImpact**

BxJ maternal	**STDEV**	**BxJ paternal**	**STDEV**
Δ14	0.018372703	Δ14	0.017525019
Δ5	0.023475713	Δ5	0.007075884
Δ12	0.032954407	Δ12	0.01102461
I1 (combined)	0.003973303	I1 (combined)	0.031955568

JxB maternal	**STDEV**	**JxB paternal**	**STDEV**
Δ14	0.002270929	Δ5	0.023018881
Δ5	0.041909731	Δ14	0.014543488
I1 (combined)	0.012850196	I1 (combined)	0.012022189
Δ12	0.023390484	Δ12	0.03021281